CHOmics Tutorial

Version 1.0, Feb, 2020

CHOmics (v1) Toolbox • My Analyses • Admin • P	rojects (z) Comparisons (zz) Sa	mplex (30)			Hello, Demo User	Ge Sign Ox
My Experiments and Analyses						E He
Vate Folder: /opt/lampp/htdocs_virtualHostings/chomics.c	rg/www.chomics.org/chomics/a M (otherwise, in FPKM)	pp_data/files_private/1000006_1kNmaZ1Noodsw1JthDnG	ANHILGBK27jAsht.pBCuDeU/			
CHO Demo	+ 🗠	RNA-Seq Data	+ 12	Test		+
2019-10-10 Samplas tsg) • Show/Hold Samples Analyses t0 • CHO Demo RNA-Seq Analysis Iv* Finished • CHO Demo RNA-Seq Analysis Iv* Finished		2019-10-08 Samples t28 + Shour/Hole Samples Analyses t0 + Shour/Hole Analyses - Test tir Prinshed		2019-13 Samples ID No samples found + Add Now Analyses found LE Start Analysis No analyses found LE Start Analysis		
My Private Projects						0 Sho
Comparisons Comparisons						C Sho
= List of Comparisons						O She

http://chomics.org

user:demo@bioinforx.com password:CHO_demo

From the login page, you can use your email to register an account that is recommended, as you will be able to save results and upload your own data. Otherwise just use guest account to view public data.

Contents

1.	Over	view of CHOmics
	1.1	Menu Bar 4
	1.2	Experiments and Analyses
	1.3	Projects
	1.4	Samples
	1.5	Comparisons
	1.6	Genes
2	Data	Input
	2.1	Upload fastq files to experiment
	2.2	Upload data file to project
3	Data	Analysis
	3.1	RNAseq analysis pipeline
	3.2	DE, GSEA and GO analysis
	3.3	Saved Genes and Comparisons 21
	3.4	Advanced Analysis
	3.3.1	Correlation Tools
	3.3.2	PCA Analysis
	3.3.3	Meta-Analysis
4	Visua	lization
	4.1	Visualize Gene Expression
	4.1.1	View Gene Expression from multiple samples
	4.1.2	View Gene Expression in Heatmap
	4.1.3	Multi-omics Expression View
	4.2	Visualize Comparison Data
	4.2.1	Dashboard View of Comparison
	4.2.2	Bubble Plot
	4.2.3	Get significant genes from comparisons 41
	4.2.4	Volcano Plot
	4.3	Visualize functional pathway 46
	4.3.1	Enrichment from Up and Down Regulated Genes 46
	4.3.2	View Changed Genes from a Functional Term in Volcano Plot

	4.3.3	View Enriched Pathways Directly from Comparison Details	. 49
	4.3.4	Multi-layer visualization	. 50
	4.3.5	Pathway Heatmap From Comparisons	. 54
5	Custo	mized analysis pipeline	. 56
	5.1	Use alternative tool or algorithm	. 56

1. Overview of CHOmics

There are several panels stacked in the main interface. The recent experiment and projects are listed in the panel separately for quick access. You can also access them and other functions from the shortcuts at the top menu bar.

CHOmics (vt) Toobox * My Analyses * Admin * Pr	gechildt Companyons hat Sample (ph)		Hallo Deno User . In Sign Cut
Welcome to BxGenomics		Shortcuts to data tables	
🗠 My Experiments and Analyses			Price
My Private Projects			
	Denela		Show/hide left menu
G All Comparisons	Panels		B Shoe
E List of Comparisons			B Shoe
● All Comparisons	Panels		Blow

1.1 Menu Bar

In top menu bar, several shortcuts are listed for quick access of functions including: Toolbox, My Analysis, and Admin, Projects, Comparisons and Samples.

'Toolbox' contains a list of functional modules including: 'Import Project Data', 'Gene Expression Analysis', 'Comparison-based Analysis', 'Pathway Visualization' and 'Other tools'.

'My Analysis' provides quick access to the information of all 'Experiments', 'Samples' and 'Analysis'.

'Admin' allows the users to manage the data files from private folder, shared folder and overview the platforms applied to all data sets.

'Projects', 'Comparisons' and 'Samples' all provide searching function and access to specific project, comparison and sample respectively.

1.2 Experiments and Analyses

Experiment is designed for running the built-in RNA sequencing pipeline on the raw sequencing data. Once the experiment is created, users can upload raw fastq files and sample meta information, and then launch the built-in pipeline for analysis. After the analysis is completed, the analysis report is generated and the results can be exported as one 'Project' for visualization and cross-project comparison.

Welcome to BxGenomics	Create Experiment for analysis	
My Experiments and Analyses	ACCOUNT (IN TAINAGE DAY)	Add sample and rur
Cruste New Experiment EC View NGS data in 1994 (otherwise, in 1994)		analysis
CHO Demo	Sample list in analysis	+ L2 Tool
	20 MPT/M0 1004 L 30104 / 500 1004 L 30104 / 500 1004 L 3017 3 L 1004 / 500 1004 L 3077 3 L 1004 / 500 1004 L 3077 3 L 1004 / 500 1004 L 3077 3 L 1004 / 500	Nes survision found di Adot Noive Analogies Bi No analopions flound del Stant Analopio
Analyses (a) + Show/Hide Analyses. + CHO Demis RNA-Seg Analysis (M Finished)	Analyses (0) Stow-Flate Analyses Tests (V* Prosted)	
	Analysis report	

1.3 **Projects**

The project is used to perform data mining and data visualization. Users can either import analysis report from 'Experiment' or upload pre-processed data to create a project. In the project, users can easily explore different features of the data (e.g, Gene expression profiling, sample clustering, PCA, differential expression genes and pathways, etc), compare the analysis with the other projects or perform the meta-analysis by combining multiple projects.

CHOmics Ival Toolbox* My Analyses * Alimin* Projects ID Companions Itali Sa	ingtin (d)	Helia Dero User 🛛 🕸 Sign Dut
Welcome to BxGenomics		
My Experiments and Analyses	Create Project	Bibw
My Private Projects		Enter
Croate New Project	Sample and comparison list	
CHO Demo - CHO Demo RNA-Seq Analysis		Rafa - Testa
Created on 2019-10-23	Created on a	519-50-83
There are 30 samples in this project.	There are a	6 samples in this project.
There are 0 comparisons in this project.	There are 4	comparisons in this project.
Analysis: (Weinight)	Analysis (9	A Tourt

Each project mainly consists of samples including both meta information and omics profiling, and comparisons showing the statistical differences among samples.

1.4 Samples

A project may include many samples which can be searched by the 'Sample' in top menu bar. Each sample has its own properties including Species, CellType, DiseaseState,etc (details available by clicking the button on the left ends).

CHOmics (v1) Toolbox * My An	alyses 👻 Admin 👻 Projects (2) Comparisons (16)	Samples (56)		Hello, Demo User	69 Sign Out
Search Sample Sear Note: You can do quick search u	Ch options Display setting	Save as sample list			
Q Advanced Search	Column Settings	eset search conditions			
Field to Searc	h Operator Value				
Age	♦ is ♦				
Search »	Add Search Condition				
Check/Uncheck All					
Column visibility Copy CSV Sh	ow 100 \$ entries			Search:	
ID	A Name	eutype SampleSource	SamplingTime	Treatment	
🖸 🗄 (C E (H (P) (1)	DK1B1D108	Transcriptomics		D108	
	DK1B2D108	Transcriptomics		D108	
🗇 🗮 (C) (E) (H) (P) (3)	DK1B3D108	Transcriptomics		D108	
	DK1B4D108	Transcriptomics		D108	
() = (C) (E) (H) (P) (S)	DK1B1D72	Transcriptomics		D72	
□ = C C (B (B (B	DK1B2D72	Transcriptomics		D72	
○	DK1B3D72	Transcriptomics		D72	
	DK1B1D84	Transcriptomics		D84	
	DK1B2D84	Transcriptomics		D84	

To change columns displayed in the table, using the table settings (green button). Users can also select the samples to save them into the sample list. Samples from the list can be loaded to other analysis or visualization stools like heatmap.

Each sample has a gene expression profile. In CHOmics, there are multiple ways to analyze and visualize the samples including: correlation tool (noted by 'C'), gene expression plot(noted by 'E'), expression heatmap (noted by 'H'), and PCA analysis (noted by 'P').

CHOmics (v1) Toolbox • My Analyses	Admin Projects (2) Comparisons (16) Samples (56)		Hello, Demo User	😝 Sign Out
Sample: DK1B1D108				
» Search All Samples				
Found in project CHO Dama, CHO Dama Phil	Can Applyin	Tools for sample analysis		
Found in project CHO Denio - CHO Denio Rick	084 D108vs D66			
Gene Expression Correlation Gene Expre	ission Plot Gene Expression Heatmap PCA Analysis			
Contraction of Contraction				
Sample Details				
ID	1	Projects ID	1	
Project Name	CHO Demo - CHO Demo RNA-Seq Analysis	Platforms ID	2	
Platform	NGS_Mouse	Platform Type	NGS	
PlatformName	Generic Mouse NGS Platform	Samples ID	27	
Species	Mouse	Name	DK1B1D108	
Description	CHO sample DK1-B1-D108, Time D108, Replicate B1	SampleIndex	27	
CellType		DiseaseCategory		
DiseaseState		Ethnicity		
Gender		Infection		
Organism		Response		
SamplePathology		SampleSource	Transcriptomics	

1.5 Comparisons

Comparison is defined by the comparative analysis between two groups of samples including differential gene analysis and pathway enrichment analysis.

There are a lot of meta data available for each comparison. See the dashboard for an overview of key categories, and the detailed description of each comparison has the full information.

CHOmics (v1) Toolbox * My Analyses * Admin * Proje	cts (2) Comparisons (16) Sample	rs (56)	Hello, Demo User 🛛 😝 Sign Out				
Search Cor Search Comparison	Display Comparison	Save to comparison list Save	to sample list				
Q. Advanced Search Change Column Settings Create Comparison List Create a Sample List * R rest search conditions Check/Uncheck All Column visibility Copy Csv Show Search:							
ID	Name	Case SampleIDs	Control SampleIDs				
2 = 8 M H C V W R K 1	D84.vs.D72	Show/Hide	Show/Hide				
	D96.vs.D72	Show/Hide	Show/Hide				
□ = 8 M H C V W R K 3	D108.vs.D72	Show/Hide	Show/Hide				
	D96.vs.D84	Show/Hide	Show/Hide				
· ≡ 8 H H C V W R K S	D108.vs.D84	Show/Hide	Show/Hide				
	D108.vs.Dg6	Show/Hide	Show/Hide				

The selected comparisons can be saved to the comparison list (yellow button) for easy loading into the plotting tools.

Several options on each comparison for complicated visualization and analysis are also listed including: bubble plot of gene expressions(noted by 'B'), meta analysis (noted by 'M'), pathway heamap plot(noted by 'H'), significant changes genes (noted by 'C'), volcano plot(noted by 'V'), Wikipathway mapping(noted by 'W'), and Rectome and KEGG pathway mapping (noted by 'R' and 'K' respectively).



1.6 Genes

The genome-wide gene expression values were detected in each sample using RNA-Seq or microarrays. All the human genes that have expression values are listed in gene table. The gene annotation from difference platforms were all mapped to NCBI gene ID (EntrezID) for consistence across platforms.

CHOmics (vt) Solbox • My Analyse	es.* Admin.* Prijechtült Companions fall	Samples (ed)			Halin, Damo Usar 🛛 🕭 Sign Gui
Search Gene +	ch box on the top-right of the table, or apply a	fvanced search before			
Q. Advanced Search 02 Change Colum	min Settings W Reset search conditions				
Column visibility Copy Coy Show	100 I entries				Search
0	* GenelAarre	Entreatio	Source	Alans	
U = (# (B) (second)	4003403.00Rk	71042	Ensembl_mouse_gene_vpa	4933401.05Rk	
11 = (10 (10 (accost)	Gma6ao6	NA.	Ensempl_mouse_gene_vga	NA	
11 = (m/m) (second)	Xir4	4020907	Ensembl.mouse.gene.vpc	AYE34250IGHI300KRG4HHI0AA388g00ir4	
1 = (1 = (1 = (ansatz)	Gmaßgeß	100418038	Ensembl_mouse_gene_vga	Grange	
U = (1) (0) (1000004)	Gmgratio	NA	Ensembl.mouse.gene.vga	NA	
10 II (10 (10 (10 (10 (10 (10 (10 (10 (10 (10	Gmg7383	NA	Enternibl, mouse, gene, vg4	NA	
0 = (#14) second	Gm32666	NA.	Ensembl_mouse_gene_vga	NA	
0 = (1) (1) (second)	Gmappa	10003/8075	Ensembl_mouse_gene_vga	Grissga'	
U = (0.0) (assessed	Gmgrgag	NA:	Ensembl_mouse_gene_vga	NA	
1) = (1)(0) (accord)	Gm/7341	664792	Ensembl.,mouse.gene.vga	EG66argalGevryan	
10 = (B) (B) (B000000)	Gregitselt	NA	Ensembl.,mouse_gene_vga	NA	
0 = 30 (0) (second)	Gmapga8	100503874	Ensembl_mouse_gene_vg4	Gmspgg8	
11 II (#) (#) (#000001	Gmaog68	100038431	Ensempl.mouse.gene.vg4	Groog88	
10 = (B) (B) (\$000000)	Gmattates	N4,	Ensembl, mouse, gene, vga	NA .	
1) = 1 (1) (100000 (1)	Grozzapli	NA.	Ensembl_mouse_gene_vg4	NA	
1 = (m. m) (200000)	Grigzgās	144	Ensembl, mouse, gene, vga	NA	
11 = 11 10 second	Gm6sos	639765	Ensempl_mouse_gene_vp4	EG6sg/8g/Gmdoot	
u = :::::::::::::::::::::::::::::::::::	Rp1	19888	EnsembL.mouse.gene.vg4	Dodc:@OrpdRoshimGue5Ros	
15 = (m. m) (account)	Gm()/483	NA	Ensembl, mouse, gene, vga	NA.	
0 = (8) (0) (second)	5047	20075	Ensempt_mouse_gene_vga	Sout/	
U = 10 (0) (0000000)	Gm3/587	NA .	Ensembl_mouse_gene_vga	NA	

To find a gene, you can use gene symbol, gene description, gene alias, NCBI gene ID, Ensembl gene ID or Uniprot ID.

For some common genes, the symbols used in publications are often not the official symbol, and you can try search alias field. For example, TP53 is often referred to as P53 in publication. You need to search P53 in alias or tumor protein p53 in description to find it if you don't know its official symbol.

The NCBI Gene search https://www.ncbi.nlm.nih.gov/gene is a good source to get official gene symbols and IDs.

You can view full details of a gene by clicking the 🗮 button .

Search All Genes	iew expression plot	View Bubble plot across comparisons		
Gene Details				
IC	0 1000003		Species	Mouse
GeneInde	x 1000003		GeneName	Gm18956
Entrezi	0 100418032		Source	Ensembl_mouse_gene_v94
Description	n predicted gene. 18956		Alias	Gm18956
Ensemb	ENSMUSG00000102851		Unigene	NA
Unipro	t NA		TranscriptNumber	1
Strand	4 +		Chromosome	1
Star	t 3252757		End	3253236
ExonLengt	h 480		GenelD	Gm18956
AccNun	n NA		Biotype	processed_pseudogene

From gene details, you can access RNA-Seq data in a box plot, or view all comparisons including this gene in a bubble plot.

2 Data Input

2.1 Upload fastq files to experiment

After the experiment is created, users can upload fastq or fastq.gz files through remote URLs, server files or local files. The files are uploaded to the private folder named 'Experiments' automatically.

CHOmics (v1) Toolbox • My Analyses • Admin • Projects (2)	Companisons (tot) - Samples (pt)	Helio, Demo User	(+ Sign Out
Experiment: Test × Delete Experiment	Manager files in folder		
Valid data files in current experiment: 0			
2. Upload Files Manago files in Private Files			
neo mus accese to one experiment yet.			
0			
Drag & Drop a File			
Or, select an option below	-11		
Permote URLs Server Files Local Fil			
and the second second			
Experiment Details @ Edit			
Description: Test			
Experiment Samples DAdd Samples One by One or In Ba	tch		
No samples found			
Experiment Analyses			

In the folder 'Experiments', there may be multiple subfolders corresponding to different experiments. Users can easily modify the folder or upload new files to the folder.

Treeview 🕮 All Files	• Se Experiments				🖿 New Falder 📱 New	File 🕹 Upleed File 🚺 Batch Up
Name	Size	Last Modified		External Link (Share among applic	cations)	Actions
S :	29.00 B	2019-10-08 11-3	Experiment folder	/Experiments/1		(X) ×
1802	1700 B	2019-10-10 091	Experiment loider	/Experiments/2	Unload files to the folder	(2°×
-	2.00 B	2019-11-20 13:28	Distance	/Experiments/3	Opload files to the folder	(2° ×

2.2 Upload data file to project

Besides raw RNA sequencing data (fastq files), CHOmics also allow the input of other types of data to start a project, including meta data (i.e, project and samples), expression data, and summary data. Those data should be uploaded in comma separated values (CSV) or tab separated values (TSV) with either

fixed or flexible format.

CHOmics (v1)	Foolbox = My Analysian = Adm	Projects (z) Comparisons (z6) Samples (56)	Hello, Derno User	EP Sign Ou
This tool will import Proj Projects -> Sampler	Import Project Data Gene Expression Analysis Gene Expression Plot Heatmap	Sion Data, and Expression Data to project sion Data. Format 2) → Sample Expression Data @ormat 2) → Comparison Data @ormat 2) → Comparison Data @ormat 2)		
Import Data Files with	Conelation Tool PC/ Analysis Export Expression Data	swith Renzie Formats		
Noter A Import	data file with fi	texture Formats Import data file with flexible format Itsvi format		
2. Select file type	Pathway Heatmap Export Comparison Data	th "Add files." button or use your computer mouse to drag and drop files into your browser. Note that, you can upload multiple files, but each file will be previewed and uploaded one by one tes, you can preset Species and Rafform. unless you have corresponding columns (Species, "Rafforms, JD, Platform, or PlatformNamel in your uploaded file. To import Sample or Compariso The amount Amount is use understored file.	a. In files, you need to select a	Project.
3. Match data file	Pathway Visualization	A ST FILMER PROVIDENT STATEMENT OF		
 To impor To impor To impor To impor Name for 	KEGG Pathway View Reactome Pathway View WikiPathway View	besides matching columns with fields, you can also set a common value for a column y which column contains Genehame, and the reat columns should match corresponding Simple Names. Ay which columns match Genehame, LogaFoldChange, PValue, and AdjustedPvalue, If you have a column matching ComparisonName, then select Matching Field "ComparisonName"; other gen, PValue, and AdjustedPvalue.	rwise, you need to enter Cor	nparison
To Impor To Impor To Impor To Impor Name for	KEGG Pathway View Reactome Pathway View WikiPathway View Other Tools My Saved Lists Functional Enrichment	s belides matching columns with fields you can also set a common value for a column. y which column contains Genehame, and the rest columns should match corresponding Sample Names. if y which columns match Genehame, LogaFoldChange, PValue, and AdjustedPValue. If you have a column matching ComparisonName, then select Matching Field "ComparisonName", other ge, PValue, and AdjustedPValue. Fixed Formats	rwise, you need to enter Car	nparison
To impor To impor To impor Name for Tips for Impor Note: All files should be	KEGG Pathway View Reactome Pathway View WikiPathway View WikiPathway View Other Toos My Saved Lists Functional Enrichment Overlap and Venn Diagrams Search Functional Gene Lists	besides matching columns with fields, you can also set a common value for a column. ywhich column contains Genehame, and the rest columns should match corresponding Sample Names hy which columns match Genehame, LogaFeidChange, PValue, and AdjustedPValue. If you have a column matching ComparisonName, then select Matching Field "ComparisonName", other ge, PValue, and AdjustedPValue. Fixed Formats mat. The first now must contain column names, which have to be the exect names of database table fields	rwise, you need to enter Car	mparison

<u>Project file</u> can be uploaded to create a new project. The project file contains some required information such as Name, Platform and other optional fields such as Disease, Description etc.

<u>Sample file</u> can be uploaded to register samples for a project. The sample file contains required information such as Name and Project_Name and optional fields such as Description, Tissue, DiseaseState, SampleSource, Gender, etc.

<u>Expression file</u> can be uploaded with quantified expression measure at gene level. The expression could be transcriptomics, proteomics or other gene-level counts. The file is required to contain GeneName, SampleName, Value.

<u>Comparison file and Comparison data file</u> are used to upload summary results for statistical comparison test applied externally. Comparison file needs to contain the Project_Name, Case_SampleIDs, and Control_SampleIDs while comparison data file contains statistical results such as GeneName, ComparisonName, Log2FoldChange, PValue, Adjusted PValue for each comparison.



3 Data Analysis

3.1 RNAseq analysis pipeline

After fastq files are uploaded to the experiment by following the Section 2.1, users can start the analysis by applying the built-in pipeline mainly including: Raw Data QC (quality control), Alignment, Gene Counts and QC, and DEG, GSEA and GO analysis. After the analysis is completed, the results can be exported into a project for visualization.

CHOmics (v1) Toolbox • My Analyses • Admin • Projects I	9 Companisons (16) Samples (56)		Helio, Demo User 59 Sign Out
Analysis CHO Demo RNA-Seq Analysis × Dela	te Analysis		
Analysis Details (r Edt Analysis Details Experiment: CHO Domo Time Created: 1003-01-0 093503 Name: CHO Domo RMA-Seq Analysis Description: (Not set) Analysis Samples: (Data Type 50) 15 samples are used. If Show All Samples: (E) Setect Samples and PR Analysis Steps and Progress	es © Censte Sample	Report	files
Tin Please select one or multiple analysis steps to per	pipeline for hitriceq		
Run Complete Analysia	Status	Step Files	
Step 1 Raw Data QC	Finished	Lef Report	
🗇 Step z Alignment with Subread 🚍	Finished	Let Report	
Step 3: Gene Counts and OC 🗮	🚺 🛩 Finished	Lef Report	
Step 4 DEG, GSEA and GO Analysis	🖥 🛩 Finished	Ed: D84vs.D72 3× GO Encement, 3× Patrixwy 3× Valwe Datal, Ed: Dg8vs.D72 3× GO Encement, 3× Patrixwy 3× Valwe Datal,	

After completion of each step, a report is generated for summarizing the metrics in each step to quantify raw data QC, alignment with Subread method, and gene count distribution and sample/gene count QC, respectively.

In the report for raw data QC, all fastq files are verified in quality by software fastQC. Sequencing read information and quality control metrics are summarized for each individual fastq file.

	BxGenomic	s - Rav	v Sequencing	Data Q	С			ר
	The fastOC program is used to verify raw data quality of the	tlumina reads				Re	ead information	
	The table below shows a summary of basic statistics. Click th	e file name to o	open the detailed report for each	ndividual sam	ple.	\subseteq		כ
Sample facto file	If the sequencing run is paired-end (PE), you may have two fi	los per sample	with R1 and R2 in the file names	respectively				
Jampie lastq lile	Filename	Total Sequences	Sequences flagged as poor quality	Sequence length	NGC	#Total Deduplicated Percentage		
	NG- 7391_790_4_RNA20140328RA_IIb44118_2577_2_1 fastq.gz	32221065	0	51	52	472%		
	NG- 7391_T108_1_RNA20140328RA_IR644119_2541_3_1/astq.gz	27636352	0	51	55	50.0%		
	NG- 7391_772_1_RNA20140328RA_Ub44108_2577_5_1 fastq.gz	47835535	0	51	53	42.7%		
	NG- 7391_TI08_4_RNA20140328RA_Ub44122_2577_2_1fastq.gz	32151429	0	61	53	46.6%		
	NG- 7391_784_1,RNA20140328RA_3044111_2541_3_1fastq.gz	25000801	0	51	53	53.3%		
	NG- 7391, T84, 4, RNA20140328RA, 8044114, 2577, 2, 1 faitig oz	29244828	0	51	52	48.7%		

he table below show pass/fail for several OC metrics. Click the file name to open individual reports. You can view fastOC documentation to get more formation about the OC metrics. Yease note that for RNA-Seq data, it is normal to observe a few failed metrics, which usually will not affect subsequent data analysis. First, per base sequence surrent fair forms content will often fail failed: due to non-random base content at the first -12 bases. This is because the random primers used during revense manuscriptors step are actually not totally random in terms of base content. Second: the sequence duplication levels of RNA-Seq data are usually high because manuscriptors.										C metrics	
File Name	Basic Statistics	Per base sequence quality	Per the sequence quality	Per sequence quality scores	Per base sequence content	Per sequence GC content	Per base N content	Sequence Length Distribution	Sequence Duplication Levels	Overrepresented sequences	Adapter Content
NG- 7391_Tg6_4_RNA20140328RA, libe4118_2577_2_1fastq.gz	PASS	PASS	WARN	PASS	FAIL	PASS	PASS	PASS	FAIL	WARN	PASS
NG- 7391_T108_1_RNA20140328RA_0b44119_2541_3_1fastq.gz	PASS	PASS	PASS	PASS	FAIL	WARN	PASS	PASS	WARN	FAIL	PASS
NG- 7391, T72, 1, RNA20140328RA, Ub44108, 2577, 5, 17ml ggz	PASS	PASS	WARN	PASS	FAIL	PASS	PASS	PASS	FAIL	PASS	PASS
NG- 7391, T108, 4, RNA20340328RA_lib44122, 2577, 2, 1 fastq gz	PASS	PASS	WARN	PASS	FAIL	PASS	PASS	PASS	FAIL	PASS	PASS
NG- 7391, T84, 1, RNA20140328RA, Ilb44111, 2541, 3, 1 failig gr	PASS	PASS	PASS	PASS	FAIL	PASS	PASS	PASS	WARN	WARN	PASS
NG- 7391, T84, 4, RNA20140328RA, Ib44114, 2577, 2, 1 fosto g2	PASS	PASS	WARN	PASS	FAIL	PASS	PASS	PASS	FAIL	WARN	PASS

In the report for alignment, parameter setting and quality metrics (e.g, mapped, junctions,etc) for alignment are listed for each fastq file.

CHOmics (v1) Toolbox * My Analyses * Admin * Projects (2) Comparisons (16) Samples (56)	Hello, Demo User 🛛 😝 Sign Out
BxGenomics - Sequence Alignment Logs	
Subread: v1.5 0-p1 (http://subread.sourceforge.net/)	
Subjunc Settings	
<pre>Function : Read alignment + Junction detection (RNA-Seq) Threads : 6 Input file : /raid/lampp/htdocs/chomics/app_data/analysis/2_yr Output file : /raid/lampp/htdocs/chomics/app_data/analysis/2_yr Index name : /var/www/html/cho_genomics/app_data/files_core/PI Phred offset : 33</pre>	
Min votes : 1 / 14 Allowed mismatch : 3 bases Max indels : 5 # of Best mapping : 1 Unique mapping : no Hamming distance : no	
Quality scores : no Summary:	
Processed : 27636352 reads Mapped : 26781244 reads (96.9%) Junctions : 111928 Indels : 46043	
Running time : 12.6 minutes	

In the report for Gene Counts and QC step, several metrics have been calculated and plotted for comprehensive evaluation of genes and samples, including: reads mapping to genes, distribution of detected genes, percentage of reads for highly expressed genes, normalization and boxplot of gene expression, sample grouping and clustering, sample correlation and outlier detection.

'1. Assign reads to genes' plots the mapping summary of reads to genes, showing the percentage of reads assigned to genes or unassigned due to unmapping, no features or ambiguous mapping.



'2. Number of Genes Detected' plots the number of expressed genes with read count from intervals of >=1, >=2, >=10, >=50 and >=100.

'3. Percentage Reads from Most Highly Expressed Genes' plots the percentage of reads mapped to the top expressed genes (up to 100 genes).

Helio, Demo User 🛛 😝 Sign Ov

2. Number of Genes Detected

Next, we performed additional QC at gene level. We first looked at number of genes detected. We count the number of genes that have at least 1, 2, 10, 50 or 100 counts. In generally, number of genes with 2 or more counts can be used as a rough estimate of how many genes are expressed. Genes with only 1 read could be noise. In addition, the number of genes with 10 or more reads is a good indicator of how many genes have enough reads for downstream statistical analysis. Click the graph to download a pdf version, you can also download a csv file containing the numbers.

Number of Expressed Genes



We also try to detect outliers from this step. Any samples that show very small number of genes with 10 or more reads are potential outliers. The cutoff we used is 1/2 of the median across all samples.

3. Percentage Reads from Most Highly Expressed Genes

We also look at the percentage of reads belonging to the top genes. Basically we rank the genes by read counts, and compute the percentage of reads belonging to the top genes (up to top 100).

If majority of the reads come from top genes, then the sample probably has bottlenecking issues where a few genes were amplified many times by PCR during library preparation. Most samples should have ~ 20% reads mapped to the top 100 genes.

If the top 100 genes account for more than 35% of all reads, we consider this sample as a potential outlier.

Click the graph to download a pdf version. You can also download a csv file download the csv file that contains the numbers.



'4. Normalization and Boxplot of Gene Expression' evaluates gene expression after normalization by TMM method and then draws boxplot of normalized expression (logCPM: log of counts per million reads) after log2 transformation.

4. Normalization and Boxplot of Gene Expression

The raw counts data were further processed by the following steps:

a) Remove genes that were not expressed. If a gene has counts per million (CPM) value >=1 in at least two of the samples, we consider it expressed in the experiment and include it for downstream QC analysis. From 32871 total genes, 14212 genes are selected as expressed and used in downstream QC analysis.

b) The TMM normalization method was used to scale samples to remove differences in the composition of the RNA population between samples. It is porformed with the ordered processor. The normalization factors for all samples	Name	group	lib.size	norm.factors
are listed below. You can download the csv file.	DK1B1D108	1	23619476	0.927
At this step, we also try to identity outliers that have extreme normalization factors (>1.5 or <0.66). Note sometimes samples with large biological differences can have extreme normalization factors.	DK1B1D72	1	41669151	1.010
	DK1B1D84	1	21522805	1.028
	DK1B1D96	1	24476888	0.995

c) The normalized gene counts were transformed to log2 scale using voom method from the R Limma package. We created boxplot for each sample to summarize gene expression.

Since this is normalized data, most samples should look similar. Samples with high or low distribution may be outliers (or have large biological differences).



Normalized Expression Values (logCPM from voom)

'5 Grouping and Clustering of Samples' plots the relationship among samples by multidimensional scale. Samples are clustered by hierarchical clustering method based on the expression of top genes with large variation(SD/mean>0.3).

5. Grouping and Clustering of Samples

a) We first create multidimensional plot to view sample relationships. This is done using R Limma package.

Here biological replicates should cluster together, and difference conditions ideally should separate from each other.

b) Very often hierarchical clustering can give better indication of the sample and gene relationships. We used made4 package from R to cluster samples and draw a heatmap.

We selected genes that have variable expression across samples to make the heatmap. These variable genes were chosen based on standard deviation (SD) of expression values larger than 30% of the mean expression values (Mean). If there are more than >5000 variable genes, we first remove genes with mean logCPM<1, then rank genes by SD/Mean to get the top 5000 genes.

The heatmap is created from 1124 variable genes.

In the heatmap above, we selected genes that changed across samples (normally by SD/mean > 0.3), and plotted the relatively gene expression levels (blue is low, red is high). Gene names are not shown due to large number of genes used to create the heatmap. Both genes and samples are clustered in the heatmap. Normally biological replicates should cluster together, and ideally there should be up- or down- regulated genes between different conditions.

Heatmap can be used to detect overall patterns, as well as outlier samples.



'6 Sample correlation' creates scatter plots for the correlation between sample pairs. The idea is that biological replicates from the same group should look similar in the scatter plot, and should have high correlation values compared to the samples from other groups.

6. Sample correlation

We also created scatter plots for the correlation between sample pairs. If there are many samples, you may need to download the graph and view it at full size. Again, the idea here is that biological replicates should look similar in the scatter plot, and should have high correlation values.



3.2 DE, GSEA and GO analysis

After completing the first three steps for sample quality control and gene count readout, users can start statistical analysis as the last step of pipeline, including differential expression analysis (DEG), gene set enrichment analysis (GSEA) and gene ontology (GO) analysis.

DEG analysis is applied to compare gene expression between two groups, namely comparison. Users can design one or multiple comparisons for DEG analysis. In each comparison, differential expressed genes are identified by LIMMA model, followed by GSEA pathway analysis and GO enrichment analysis which explore the enrichment of DEGs in diverse pathways.

CHOmics (vt) Tostox • Hy Anie)	nes* Amin* Prair	da lati Companioni b	di Samplan (60)			Helis Denis Unit - M Sign Gu
Analysis Test XOMM Avagues						
Analysis Details grade Analysis Details	ysis Detats					
Time Created: 2032-13-22 08:50:10 Name: Test Description: (Not set)						
Analysis Samples Data Typ	w PD					
15 samples are used # Show All Samples	s 18 Select Simples in	d Flass O Create Som	pla			
Analysis Steps and Progr	ress					
Dupticity Analysis	Alyon Brovew Fal	EReport				
Tip: Please select one or multiple analysis r	stops to get started					
Run Compiletia Analysia					Status	Step Film
U Step t Raw Data OC					8 of Destand	Let Report
Step z Alignment with Si Step is Create County and	ubread III			Parameter for ge	no set analysis	Lef Report
K Step & DEG, OSEA and G	O Analysia			Farameter for get	ile set analysis	
DEG Parameters				7		
Use TMM Normalization	True					
Minimum number of Genes II a set.	н					
Maximum number of Genes in a set.	1000				Design comparison for D	EG analysis
	084		vs 072	•		
	Dgli	×.	vii 072	• *		
Salart Comparison	096		vs 079	· • 😠		
	096		vs 084	• *		
	D108		vs 084	• *		
	Doold	•	vs Dg0	• *		
	Il Add Comparison					
(Section of the sect						

The reports for DEG and pathway analysis are attached for each comparison after completion of analysis.



In the report for DEG analysis, the table summarizing DEGs with up- and down-regulation is listed along with a heatmap clustering the DEGs expression (up to top 1000 DEGs).

OLLICE MT 20000K. WAYONIAME. Value, Note	tar will Companyona davi Samples (96)			Holio, Demo Usor - 84 Sig
	Comparisons D44+xDy2 Dg8/	s.Dyz Doollivs.Dyz Dyślivs.Dite Doollivs.Dite	Cholivs.Dgd	
	Diff		C Duck Day	
	Diffe	rentially Expressed Gen	es for D108.vs.D72	
	1. Summary of Differentia	lly Expressed Genes (DEGs)		
	For comparison, the number of different regulated genes are high in the second	taily expressed genes (DEGs) are shown below. Up- condition 0	political contraction if while down-	ר
	Cutoff	Standard Itwo fold and FDR 0.051	DEGs summary table	
	# of Up-regulated genes	171	100	, ,
	# of Down-regulated genes	46	54	
	Total number of changed genes	216'	411	
	These genes are reported in the table	for differentially expressed genes (DEGs) and shown i	n the DEG heat maps.	
	Summary heat map showing all the diff	erentially expressed denes (DECs)	Calue Kay	
	sampide, Sanh tervi se gene debri dol shown in this plot. Click image to down if View detailed heat may browing nar soop Click, only the tes stoop genes in Sometimes, you may want to view gene Lampide debri unchanged. Nou can view	Used and voted of County and the office of the office office of the office offi	DEGs Expression heatmap)
	Decision Proc.			

In the report for GO enrichment analysis, barplots show the significance of enrichment of up- or downregulated DEGs in different pathway databases, e.g, GO, KEGG, Wiki pathways, etc.

Ge	ene Ontology Enrichme	ent Analysis Resu	ults		
Introduction This page displays the top 10 lists	from functione; enrichment of differentiali expresse	Pathway a	nalvsis for up-	regulated DEGs	
Comparisons DH4vsD72	95/vs.D72 Diollivs.D72 Dg6/vs.D14 Diolliv 19 Dourrisoid SVG File	17AU			
Up Regulated Genes					
Rictograt Process		Biological Process			
Cebuar Component	positive regular	ion of contractions secretion	log(2)-4.34		
Molecular Function	regular	ten of contracting secretion	August 4 DE		
RECO	Negative regulation to	Firmunopobulin production	10(2)-4 (0)		
Molecular Signature	positive regulation of regulation of	immature T cell proliferation	49231-8-30 Avg321-8-21		
Interpro Protein Domain	positive repulation o repulatio	f striated muscle contraction	ingipt-7.75 ingipt-7.37		
Wild Pathway	repartive reputation of mu	prozein refelding	kgp)-7.24	Aug21-7.18	
Reactome					
W Enrichment Report		Pathway a	nalysis for dov	wn-regulated DEGs	
Down Regulated Genes	B Download SVG File				
Rickopick Process		Biological Process	10 CL 14 17 10		
Celàular Component		-			
Molecular Function		response to hypoxia	Kog(2):-10.19		
KEDG	- response regative reputation of	e to decreased oxygen levels	krg(3)-8-87 mg(3)-8-24		
Molecular Signature		response to nitric oxide	log(s)-0.10 log(s)-0.00		
Interpro Protein Domain		rysnic platenta development	huggis-it at		
Wiki Pathway	cellular response response to a	wygen containing compound		N9010-7.90	
Reactorne		response to dexamethasone	2 4 6		
			Number of Genes		

Similarly, in the report for GSEA analysis, enrichment results for up- and down- regulated DEGs in pathways from MigDB database are plotted with significance level (FDR), respectively.

CHOMICS (V1) Toolbox * My Analysis * Acres * Projecta III Comparisons Intil Samples	satu		Hello Demo User 🛛 😻 Sign Out
	BxGenomics - Su	ummary of GSEA Analysis	
late	aduation		
1111C	Juduction		
This p	age displays the top 10 lists from functional enrichment of differ	rental expressed genes:	
10 GS	EA Topso Genes Summary Data for All Comparisons (CSV form	it, open in Excel)	
D10	8.VS.D72 IN View Detailed Report		
		合约 表示 推进 经正确 把 算	
	Negative Enrichment	Positive Enclohment	
	DNA REPLICATION WIRL PORCO DO	NASA ECH SCHOPROTEL	
	STATISTISS OF DIA KD	EXTRACLUCAR MATRIX	
	FCK1 SOURCEPU EVINI.	COLLAGEN CONTAINING	
		AND MADE PROPERTY. INC. CONTRACT, CONTRA	
	AND IN AT REFUSANCE AND A DECK AN	THE ADDRESS OF THE F	
6	The STRAND & Goldania Strand	PATRACTURE AS MATOR	
		And Constanting and State	
	DNA DEPENDENT DNA BE	EVERATE LINE AN INVERSE TO THE POST OF	
	UN I MARY AND TARGETS	ANA WITCHING MANY	
	0 50 500 500	8 100 200 300	
	Number of Cenas	Number of Finner	

3.3 Saved Genes and Comparisons

Customers can save selected genes or comparison for future use (e.g. multiple gene and multiple comparisons bubble plot). From gene search, check the genes you want to save, and click the yellow button "save selected genes

CHOmics (v1) Toolbox • N	Ay Analysies * Admin * Projects	(1) Comparisons (12)	Samples (30)		Hello, Demo User 🛛 😣 Sign Out
Search Gene Note: You can do quick search using	e a the search box on the top-right o	f the table, or apply a	tvanced search below.		
Q, Advanced Search Q; Char Check/Uncheck All Column visibility Copy CSV	nge Column Settings	s Gene List >> Reso	3. Save	1.	Search Term
D	* GeneName	EntreziD	Source	Alias	2
😰 🗏 (1) (2) (1000000)	glutamate	57	NA	NA	
🛛 🖽 (B) (G) (11000021)	elutamine	53	NA	NA	
🛛 🗮 (B) (G) (1000022)	2. Check gene	s of interest	s A	NA	
🗇 🗮 📵 🕼 (1000023)			in .	NA	
(I)	N-acetylglutamine	33943	NA	NA	
🗇 🗏 🛞 🌀 (1000005)	glutamate_gamma- methylester	33487	NA	NA	
🗇 🗮 (B) (G) (1005026	pyroglutamine	46225	NA	NA	
(B) (G) (B000027)	N-acetyl-aspartyl- glutamate_NAAG	35665	NA	NA	

You can do the same with comparisons.

CHOmics (v1) Tootbox • My Analyses • Admin •	Projects (1) Comparisons (12) Samples (30)		Hello, Demo User 🛛 🚱 Sign O
Search comparison +	a side of the state or marking an end of state builds		
Advanced Search Charge Column Settings Check/Uncheck All	Pright is the label of apply actes of a meet for any it. Pright is the label of apply actes of a meet for any it. Pright is the label of a meet for a meet for any it. Pright is the label of a m	tions	Search Term
Column visibility Copy CSV Show 100 \$	Save comparison 4. Save sample list	:]	Search: D108
ID	* Name	Case SampleIDs	Control SampleIDs
	D108.vs.D72	Show/Hide	Show/Hide
	D108vsD84	Show/Hide	Show/Hide
O = e M H C V M B K I		Show/Hide	Show/Hide
2. Chec	k genes of interests	Show/Hide Show/Hide	Show/Hide Show/Hide
	k genes of interests Comparison, Protein, Dzoß, D72	Show/Hide Show/Hide Show/Hide	Show/Hide Show/Hide Show/Hide

To view selected genes or comparison, click "My Results" link on the left menu and then "Gene Lists".

e. You can do quie	Gene Expression Analysis	top-right of the table, or apply	advanced search below.		
	Gene Expression Plot				
dvanced Searc	Heatmap	Create Comparison List	Create a Sample List PReset search conditions		
ck/Uncheck A	Correlation Tool PCA Analysis				
m visibility	Export Expression Data	ies.			
	Comparison-based Analysis		Name		
	Volcano Plot		DinRue D23	-	
	Bubble Plot Significantly Changed Gener		DEDRUG DRA		
	Pathway Heatmap		010035004		
(B) (M) (H) (I	 Export Comparison Data 		D108 vs.D96		
00 00 00 0	Pathway Visualization		Comparison_Protein_D108_D72		
E (B) (M) (H) (I	KEGG Pathway View		Comparison_Protein_D108_D84		
	Reactome Pathway View		Comparison_Protein_D108_Dg6		
g 1 to 6 of 6 er	 WikiPathway View 				
	Other Tools				
	 My Saved Lists 				
	My Saved Lists Functional Enrichment	Click to swit	ch to different saved lists		
	My Saved Lists Functional Enrichment Ounclan and Mona Diagrams	Click to swit	ch to different saved lists		
	My Saved Lists Functional Enrichment Overlap and Venn Diagrams	Click to swit	ch to different saved lists		
	My Saved Lists Functional Enrichment Overlap and Venn Diagrams Search Functional Gene Lists	Click to swit	ch to different saved lists		
	My Saved Lists Functional Enrichment Overlap and Venn Diagrams Search Functional Gene Lists Compare Gene Lists	Click to swit	ch to different saved lists		
	My Saved Lists Functional Enrichment Overlap and Venn Diagrams Search Functional Gene Lists Compare Gene Lists Meta Analysis	Click to swit	ch to different saved lists		
	My Saved Lists Functional Enrichment Overlap and Venn Diagrams Search Functional Gene Lists Compare Gene Lists Meta Analysis Manage Platforms	Click to swit	ch to different saved lists		
	My Saved Lists Furctional Enrichment Overlap and Venn Diagrams Search Functional Gene Lists Compare Gene Lists Meta Analysis Manage Platforms	Click to swit	ch to different saved lists		
IOmics (v1)	My Saved List Functional Enrichment Overlap and Venn Diagrams Gearch Functional Gene Lists Gompare Gene Lists Meta Analysis Manage Platforms Toolbox * My Analyses*	Click to swit	risons (12) - Samples (30)		Hollo, Demo User 🔗
Omics (v1)	My Saved Lits Functional Enrichment Overlap and Venn Diagrams Generh Functional Gene Lists Gompare Gene Lists Meta Analysis Manage Platforms Toolbox * My Analyses * /	Click to swit	ch to different saved lists		Hello, Demo User 🔗 🖗
Omics (v1) Saved	My Saved Lits Functional Enrichment Overlap and Venn Diagrams Generh Functional Gene Lists Gompare Gene Lists Meta Analysis Manage Platforms Toolbox * My Analyses * Lists	Click to swit	ch to different saved lists		Hello, Demo User 🔗 6
(Omics (v1) Saved	My Saved Lits Functional Enrichment Overlap and Venn Diagrams Search Functional Gene Lists Gompare Gene Lists Meta Analysis Manage Platforms Toolbox * My Analyses * Lists Overlap Analyses * Overlap Analyses	Click to swit	ch to different saved lists	Count	Hello. Demo User 6
IOmics (v1) Saved	My Saved Lits Functional Enrichment Overlap and Venn Diagrams Generh Functional Cene Lists Gompare Gene Lists Meta Analysis Manage Platforms Toolbox * My Analyses * / Lists G % All saved G % All saved	Click to swit	ch to different saved lists	Count i	Hello, Damo User 6
HOmics (v1) V Saved Jory: (ALD Jumn visibility	My Saved Lits Functional Enrichment Overlap and Venn Diagrams Generh Functional Gene Lists Gompare Gene Lists Meta Analysis Manage Platforms Toolbox * My Analyses * / Lists @ ** All coved Copy Csy Show 100	Click to swit	ch to different saved lists	Count i	Hello: Domo User 6
Omics (v1) Saved ory: (All) me	My Saved Lits Functional Enrichment Overlap and Venn Diagrams Generh Functional Cene Lists Gompare Gene Lists Meta Analysis Manage Platforms Toolbox * My Analyses * / Lists G % All soved Gay Csy Show 100	Click to swit	ch to different saved lists risons (ta) Samples (go) Category	Count	Hello: Domo User 6 in list Search
Omics (v1) Saved mr visibility me ele_demo	My Saved Lits Functional Enrichment Overlap and Venn Diagrams Search Functional Cene Lists Compare Gene Lists Manage Platforms Toolbox * My Analyses * / Lists @ WALcoved Copy Csv Show 100 x	Click to swit	ch to different saved lists risons (ta) Samples (30) Comparison Comparison	Count i G	Hello, Demo User 0 in list Search: Time 2020-02-25
Omics (v1) Saved ary: (All) mr visibility me ole_demo	My Saved Lits Functional Enrichment Overlap and Venn Diagrams General Arthonous General Arthonous General Arthonous General Arthonous Manage Platforms Toolbox O	Click to swit	ch to different saved lists risons (tz) Samples (307 Category Comparison Gene	Count i 6 15	Holio, Demo User 0 in list Search: Time 2020-02-25 2020-02-25
Omics (vt) Saved my visibility me vise_demo	My Saved Lits Functional Enrichment Overlap and Venn Diagrams Search Functional Gene Lists Orangae Gene Lists Manage Platforms Toolbox * My Analyses * 1 Lists O * All caveo Copy Csv Show 100 List List	Click to swit	ch to different saved lists risons (12) Samples (30) Category Comparison Gene Gene Gene Gene	Count i 6 15 169	Hello. Demo User 0 in list Search: 2020-02-25 2020-02-25 2020-02-25 2020-02-25
Omics (v1) Saved ary: (All) m visbity ree ke_demo s_bio8vs.D7. rs_allcompar	My Saved Lits Functional Enrichment Overlap and Venn Diagrams Search Functional Gene Lists Orangae Gene Lists Manage Platforms Toolbox My Analyses Verta Analysis Manage Platforms Lists Overlap Analyses Lists Lists	Click to swit	Chito different saved lists risons (12) Samples (30) Category Category Comparison Gene Gene Gene Gene Gene Gene Gene Gene Gene	Count i 6 15 169 19	Holo. Demo User 0 in list Search 2020-02-25 2020-02-25 2019-10-24 2020-01-14

One additional way to select and save genes or comparisons is from the 'significant changed genes' in toolbox. Using the dynamic filters to choose the comparisons or genes you are interested in, and you can use the table at the bottom to save comparisons or genes.

CHOmics (v1)	Toolbox • My Analyses • Admi	n * Projects (J)	Companisons (12) Samples (30)		CHOmics (v1) To	polbox * My An	alyses * Admin	 Projects (1) Comparisons 	2. load comparisons from project
Search Note: You can do quie	Import Project Data Gene Expression Analysis Gene Expression Plot	top-right of th	he table, or apply advanced search below.		Significantly	Change	d Genes		
Q, Advanced Searc	Heatmap	Create C	Comparison List Create a Sample List >> Res	set search conditions	Comparisons: » Load	from saved lists	Q Search and	Select Q Select a Project 🗴	Clear
Check/Uncheck Al	Correlation Tool PCA Analysis				D84.vs.D72				
Column visibility	Export Expression Data	ies	1 aliak alamifiaanti	webeneed serves	D96vs.D72				
0 0 = 0 0 0 0	Comparison-based Analysis • Volcano Piot • Bubble Piot		1. Click significanti	y changed genes	I				
	Significantly Changed Genes		Dt08.vs.D84		L				3. Filter genes
	Pathway Heatmap Except Comparison Data		Dto8vs.Dg6		Or, upload your comparie	ison files: Choos	se File No file ch	osen	Si i ittel genes
	- coport comparison costs		Comparison_Protein_D108_D72						
○ ■ (8) (8) (8) (8)	Pathway Voualization KEGG Pathway View		Comparison_Protein_D108_D84		Display Options: 📴 Log	g2FC 🖸 P.Value	FDR		
	Reactome Pathway View		Comparison_Protein_D108_Dg6		Fold Change Cutoff: a	z	e Both Up-a	nd Down-regulated	
Showing 1 to 6 of 6 er	 WikiPathway View 								
	Other Tools • My Saved Lists				Statistic Cutoff: FDR	• ≤	0.05 0		
	Functional Enrichment Overlap and Venn Diagrams				List Genes: O Common	in Genes from Al	Comparisons	Genes from Any Comparisons	
	Search Functional Gene Lists				Submit Start Ov	er -			
	Compare Gene Lists				Constant of the local division of the local				
	Meta Analysis Managa Blatforms								

Subr	Start Ov	4. Save comparison list									
• Sav	e Comparison List	Bubble Plot Pathway Heatmap Meta Analysis Expore	t Comparison Data	WikiPathways	Reactome Pathwa	iys 🕨 KEGG Pathy	nays				
+ Sav	e Gene List 🔹 G	ene Expression Plot + Heatmap + Correlation Tool + PCA Analy	sis + Export Expres	sion Data							
Colun	nn visibility Co	6. Save gene list								Search:	
	GeneName	A Description	D84.vs.D72 - Log2FC	D84.vs.D72 - P.Value	D84.vs.D72 - FDR	D96.vs.D72 - Log2FC	Dg6.vs.D72 - P.Value	D96.vs.D72 - FDR	D108.vs.D72 - LogzFC	D108.vs.D72 - P.Value	D108.vs.D7 - FDR
B	Adm	5. Select genes	-1.0525	0.0000	0.0008	-1.3996	0.0000	0.0001	-13009	0.0000	0.0022
	Aqps		-1.4112	0.0000	0.0000	-16577	0.0000	0.0000	-1.5824	0.0000	0.0000
	Arhgef37	Rho guanine nucleotide exchange factor (GEF) 37	-12755	0.0000	0.0024	-1.3097	00000	0.0002	-16978	0.0000	5000.0
	Calor	calcitonin receptor	1.3005	0.0001	0.0071	1.9313	0.0000	0.0001	2.2895	0.0000	0.0000
	Capn8	calpain 8	11739	5000.0	0.0088	1.4216	0.0000	0.0011	11208	0.0016	0.0204
	Cdh17	cadherin 17	2 1087	8000.0	0.0192	2.6875	0.0000	0.0009	2.8212	00000	0.0006
	Cdknac	cyclin-dependent kinase inhibitor 1C (P57)	-1.1140	0.0001	0.0051	-1.5953	0.0000	0.0005	-14709	0.0000	0.0010
	Celf5	CUGBP. Elav-like family member 5	1.4453	0.0006	0.0173	1.4787	0.0015	0.0140	11600	0.0059	0.0479
	Chad	chondroadherin	12709	0.0001	0.0067	1.7458	0.0000	0.0003	17059	0.0001	0.0024
	Crisptd2	cysteine-rich secretory protein LCCL domain containing 2	18740	0.0023	0.0345	13480	0.0022	0.0181	19423	0.0004	0.0087

3.4 Advanced Analysis

Besides the above analyses, the CHOmics also provides several advanced tools.

3.3.1 Correlation Tools

Once the user has identified a gene of interest, the user can use correlation tools to find other genes that share similar (or opposite) profiles in terms of gene expression or fold change. First, enter the gene of interest, and samples to be used for correlation. In the example below, we entered a saved gene list, and 15 samples.

and the second se	tort horo				Helia, Demo User - 14 Sign Out
Correlation Tool					
Advertish Advert	gene of ests		Sergies: ** Coal from savet list: @ Generic and Se Coal Door Coal Door Coa	nd Glassishipet xDee r load nples	
Advanced Options	4. Set options]			
Calculate the correlations against all availate	the genes in database * Calculate the correlations among the enter	red genes only:			
Correlation Method:					
# Pearson Correlation 10 Spearman Correla	ion Pearson Correlation Coefficient Between Ranked Variables/				
Enable Log, Transform	_				
# Submit Higher					
supe records found a	——/ The result show	ws a table of correlated g	enes ranked by R ²		
			circs, runkeu by h		
				J	
Column visibility Copy CSV Show 30	a \$jentres			J	Search
Column visibility Copy CDV Show 30 Source Cene	a B entres Matched Gane	Correlation Coef	· RSquare	# of Data Points	Search Actions
Column valuation Colory Color Show 10 Source Gene	o B jertines Mutched Gene Cita	Correlation Coef	• R Square 0.0534	er of Data Points 15	Search Address Address Ed Place
Cosum vability Cosy Cosy Show 10 Source Come	o 1) entries Matched Gane Clas Clab	Correlation Coef o gata o gata	R Square 0.0034 0.0034	# of Data Points 25 25	Search Actions Let Pluk Let Pluk
Column validative Copy Copy Street 10 Source Gene Cont Cost Grin	B instead Mitched Game Clais Clais Clais Mitched Mittige	Constation Cost 0.9957 0.9968 0.9968	R Square 0.0034 0.0004 0.0004 0.0004	H of Data Points 25 35 35 35	Skarch Address Address Mit Pluck Mit Pluck Mit Pluck
Colum Habity Coay Show 19 Source Gene Cost Cost Cost Sergents	Mathed Gare Clair Clob Mispo Clair	Constation Coal . 0.9077 0.9546 0.9996 0.9971	R Square 0.0034 0.0034 0.0031 0.0031 0.0131 0.0137	# of Data Parts 15 15 15 15 15	Search Actions ME Rick ME Rick ME Rick ME Rick ME Rick
Courre Healing Copy Copy Show 19 Source Game Chil Chil Chil Chil Chil	Burnes Mitched Gene Cita Cita Cita Cita Cita Cita Spit	Constation Coef . 0.0545 0.0545 0.0545 0.0545 0.0545 0.0545 0.0545 0.0545 0.0545 0.0545 0.0545 0.0545 0.0545 0.054 0.054 0.054 0.054 0.054 0.054 0.054 0.054 0.054 0.054 0.054 0.054 0.054 0.054 0.0545 0.05555 0.05555 0.05555 0.05555 0.05555 0.05555 0.05555 0.05555 0.05555	Rispure Oppid	# of Data Norts 16 15 15 15 15 15 15	Search Actions Ind Proc Mel Proc Mel Proc Mel Proc Mel Proc Mel Proc Mel Proc
Courre valation (Coary) Coar Show (So Source forme Cost Cost Cost Cost Cost Cost Cost Cost	B proteins Mitched Game Cita	Constation Cold 0.005 0005 0005 0005 0005 0005 0005 0005 000	Rispure oppie oppie oppie ophie	H of Data Parts 55 55 55 55 55 55 55	Skarch Addown Addown Addown Add Nac Add Nac Add Nac Add Nac Add Nac Add Nac Add Nac Add Nac
Column Walking Copy Copy Show 100 Source Game Coni Coni Coni Simprifi Colini Simprifi UccConflogoto	Matched Game Cital Cital Cital Cital Cital Grad Spon Spon Spon LOCisory7888	Constation Coal 0.0ptr/ 0.0pe8 0.0pe8 0.0pe8 0.0pe8 0.0pe8 0.0pe8 0.0pe8 0.0pe8 0.0pe8	Rigum Oppid O	ar of Data Punts 15 15 15 15 15 15 15 15 15 15	Search Addone Addone Ed Nuc Ed Nuc Ed Nuc Ed Nuc Ed Nuc Ed Nuc Ed Nuc Ed Nuc
Course Gane Gane Gan	Butched Game Mitched Game Clus Cloir Mings Cloir Son Son Son Son Son On	Constation Cold 0.564 0.564 0.564 0.565 0.	R Square Ogold	# of Data Points 15 15 15 15 15 15 15 15 15 15	Search Actions Ed Nuc Ed Nuc
Course Venters Source Gene Crist Crist Sergents Colle Sergents UCEODOfoggie Sergents Colle	Burnow Matched Gene Cisis Cisis Cisis Cisis Cisis Cisis Sen Cisis Cisis Cisis Cisis Cisis Cisis Sen Cisis Sen Cisis Cisis Cisis Cisis Sen Cisis	Constation Cost - 0.0947 0.0948 0.0945 0.0945 0.0954 0.0954 0.0954 0.0954 0.0954 0.0954 0.0954	Require contract by relationships contract	H of Data Noiels 25 36 45 55 56 55 56 55 56 55 56 55 56 55 56 56 56 56 56 56 56 56 56 56 56 56 56 56 56 56 56	Search Antona Antona Kel Murk Kel Murk Kel Murk Kel Murk Kel Murk Kel Murk Kel Murk Kel Murk Kel Murk
Course Version Course C	I entres Mitched Dane Ctal Ctal Ctal Ctal Ctal Sen Sen LOCson7785a Gn Sope Ctal	Consistion Cold 0.9957 0.9968 0.9968 0.9968 0.9968 0.9968 0.9968 0.9968 0.9968 0.9969 0.9969		If of Data Paris 85 86 85 86	Search Actions
Course Series Source Series Coin Coin Coin Coin Coin Coin Coin Coin	Matched Game Cita Cita Cita Cita Cita Cita Soria	Consistion Cost 0.9017 0.9040 0.9040 0.9041 0.9043 0.9044 0.9044 0.9044 0.9044 0.9044 0.9049 0.9049 0.9049 0.9049 0.9049 0.9049 0.9049 0.9049	R Square 0.0034 0.004 0.004 0.004 0.004 0.0057 0.0059 0.0059 0.0059 0.0059 0.0059 0.0059 0.0059 0.0059 0.0059 0.0059 0.0059 0.0059 0.0059 0.0059 0.0059 0.0077	If of Data Paries 85	Search Actions Edition
Course Gener Source Gener Crist Crist Crist Sergen 5 Color Sergen 7 Color Sergen 7 Color Sergen 7 Color Sergen 7 Color C	3 entres Matched Gene Cis Cis Cis Cis Sen Sen Sen Sen Sen Cis Sen Cis Sen Cis Sen Se	Constation Cost / 0.0047 0.0048 0.0	Rispure 0 0934 0 0934 0 0935 0 0937 0 0937 0 0937 0 0939 0 0939 0 0939 0 0939 0 0939 0 0939 0 0939 0 0939 0 0939 0 0939	He of Data Noiets 25 36 45 35 45 36 36 35 36 36 36 36 36 36 36 36 36 36 36 36 36 36 36	Search
Course Name In Course Topology	I entres Mitched Gene Cisle Cisle Grafi	Constants Cold 0.957 0.958 0.958 0.958 0.958 0.958 0.958 0.958 0.958 0.958 0.958 0.958 0.958 0.958 0.959 0.959 0.959		If of Data Paints 85 86 85	Search Addow Edition E



Click the plot icon will show scatter plot of the target and the correlated gene (e.g, gene Grn vs. Ctsa).

3.3.2 PCA Analysis

You can select a set of samples and genes and use PCA plot to visualize the sample relationships on the target gene set.

PCA Analysis	
Saved Results + PCA tack for uploaded data files Genes => Load from saved lists =Q Load functional game sets =x Clear gene of interests	S Samples: » Load from saved lists: Q Search and Select: Q Select a Project. * Clear saved samples
Adamta6 Adamta6 Adamta7 Actor Arpa1 Arpa3 Arpa3 Select sample attributes for visualization	DistalsDate Distal
Sample Attributes: Sengle Attributes: Age Colltype Collection DeceseCategory DecessState Ethnicity Flag Remark Flag To Response SampleRuthology SampleSource SampleType SamplingTime Symptom Tissue Submit	Remove Gender Infection Organism RN Number RNASeq Assignment Rate RNASeq Mapping Rate RNASeq Total Read Count TasseCategory Transfection Transfection Transfection Transfection Count Uberon ID Uberon Term Counce All Counce

The system will use FactorMineR package to run PCA analysis and display the results. Several PCA metrics are plotted for interactive visualization:

<u>Eigenvalues</u> plot the percentage of variance explained by top PCs.

Variables Plot shows the weights of top contributing genes in each PCs.

Variable Data summarizes the weights of each gene in each individual PC.

Individuals Plot shows the relationship of samples on the spanned space by different PCs.

Individual Data summarizes the score vector of each sample in each individual PC.





The PCA results can be saved. Users can load it in the future.

My PCA Saved Results

Genes & Samples Analysis Saved Results PCA tool for uploaded data files		
Show 300 \$ entries		Search:
Title	* Description	Actions
Test_PCA		# Delata
Showing 1 to 1 of 1 entries		Previous 1 Next

Users can upload their own data matrix or pre-calculated data for PCA analysis and visualization.

PCA Scat	ter Plot To	ool	
Genes & Samples Ana	lysis + Saved Results +	PCA tool for uploaded data files	
Upload Files: Da	ta file is required		
	Data File:	Choose File No file chosen	Example Data File
Data matrix for PCA	Attributes File:	Choose File No file chosen	Example Attributes File
)	Variance File:	Choose File No file chosen	Example Variance File
	Format	ocsv ⊖txt / tsv	

3.3.3 Meta-Analysis

Meta-Analysis can be used to identify genes that are changed consistently across multiple projects. It is listed as one functional module in toolbox panel. In the example below, we are looking for the most significant DEGs in three comparisons.

eta Anatysis	Save meta-analysis	Enter or load comparison list	
es: » Load from saved lists Q Load	functional gene sets X Clear	Comparisons » Load from saved lists Q Search and Select Q Select a	Project X Clear
idamtső idamts7 Kdhs Enter or load ge	ene list	D64.vs.D72 Dg63vs.D84 D108.vs.D96	
ebp1 ff2 mpd3 ne Attributes: (3 selected) >> Show A	Attri	ibutes to show in results table	
Vit2 Vit2 Vit2 Vit2 Vit2 Vit2 Vit2 Vit2	tributes Attri	ibutes to show in results table	lone
Vit2 vit2	ttributes	ibutes to show in results table	lone
ebps #2 mpd3 re Attributes: (3 selected) >> Show A SeneName @ EntreziD Source @ De vanced Settings: >> Toggle Masing data allowed for P-value: Log2FC cutoff (s is 2 fold);	tributes	ibutes to show in results table	lone
ebp: #2 mpd3 ne Attributes: {3 selected } > Show A GeneName © EntrealD Source © De vanced Settings >> Toggle Missing data allowed for P-value: LogaFC cutoff (s is 2 fold): Statistical type for changed gene:	tributes scription Alias Ensembl	ibutes to show in results table Unigene Uniprot TranscriptNumber Strand Chromosome Start End ExonLength AccNum Check All XCheck N Optional parameter setting for cutoff on meta-analysis statistical results	lone

The meta-analysis pipeline will compute three types of results:

- Maximum p-value (maxP). This method targets on DEGs have small *p*-values in "all" comparisons. We recommend using maxP if you are looking for DEGs that are common among several studies.
- 2) Fisher's p-value. The Fisher's method sums up the log-transformed p-values obtained from individual studies. This p-value combination method is useful if you want to identify DEGs in any of the comparisons.
- 3) We also applied simple counting method to report the frequency a gene is classified as up or down-regulated DEG from all the comparisons. The default DEG cutoff is two-fold change and FDR<0.05. but user can change the cutoff.</p>

In most cases, combing maxP (smaller values are more significant) and the counting method (e.g. upregulated in 50% of studies) will give the most biological relevant results for consistently regulated genes across comparisons.

	w Bubble Plot	a Save Genes	& Downlo	oad Meta Data	B Save Result	s »F	iter Results	Filtering	Colur	nns in results tabl	e
oggle Up Co D8 D1	columns: Per mbined_Pval_Fishe 4.vs.D72_FDR 98.vs.D96_FDR	r æ	GeneName Down.Per Combined_ Dg6.vs.D84	Pval_maxP _logFC	 ✓ Entre ✓ Rank ✓ Coml □ D96x 	zID Prod bined_FDR_Fit bined_PVal	sher	Description RP_logFC COmbined_FDR_maxP D96vs.D84_FDR	© GeneIndex ⊘ RP_Pval © D84vs.D72.log © D108vs.D96_b	■ N.da RP_F D84v pgFC ■ D108	ta points IDR vs.D72_Pval vvs.D96_Pval
Colur	nn visibility Copy	CSV She	ow 25 🗘	entries						Sec	arch:
6	GeneName	EntreziD	Up.Per	RankProd	RP_logFC	RP_Pval	RP_FDR	Combined_Pval_Fisher	Combined_Pval_maxP	Combined_FDR_Fisher	Combined_FDR_m
8	Mb21d1	214763	66.6667	8.7760	0.8800	0.0033	0.2758	0.0000	0.0000	0.0000	0.0000
8	Snord14e	100302594	33 3333	7.9260	1.0500	0.0023	0.3927	0.0000	0.0000	0.0000	0.0000
8	Hspata	193740	33-3333	15.7900	0.8400	0.0199	0.6728	0.0001	0.0001	0.0002	0.0004
Ð.	LOC100689269	100689269	33.3333	14.9500	0,7800	0.0170	0,7182	0.0000	0.0000	0.0000	0.0000
	LOC100689270	100689270	33 3333	13.9000	07400	0.0138	0.7744	0.0000	0.0000	0.0000	0.0000
107	Calcr	12311	33 3333	36.0400	0.7500	0.1531	1.8350	0.0000	0.0000	0.0000	0.0000
e											

In the above example, we used a relatively loose filtering criterion (N.data.points>1, and up-regulation in percentage>30% of studies, and Combined_Pval_MaxP <=0.0001) because only small number of genes pass the stringent default criterion.

DI COTO DOIDIC S- 1		
IN.data.points >= 1		
Percentage Up-Regulated	30	
Percentage Down-Regulat	ted >- 0.01	
Combined_Pval_MaxP Cut	toff <= 0.0001	
RP_Pval <= 0.01		
Pval Fisher Cutoff <= 0.02	1	
lumber of records to show:	● 100 ◎ 1000 ◎ 3000 (limit)	

The data table shows the genes that pass the filters. We can sort the table by maxP value. A different filter can be applied to get down-regulated genes.

The results can be saved for future access. There are also links to several other tools. The download meta data link will save a CSV file that contain results from all genes.

Next, we will choose all the genes that pass filter by checking the box for all listed genes, and use bubble plot to visualize the results.

Bubble Plot Multiple	
» Single Gene Plot	
Genes: » Load from saved lists Q Load functional gene sets X Clear	Comparisons: » Load from saved lists Q Search and Select Q Select a Project X Clear
Calcr Hispata LOC:00689270 LOC:103832837 Mo2:dt Note: You must enter one or more gene names.	Dis08vs/Dg6 DB4vs/Dg2 Dg8vs/D84 Note: You must enter one or more comparison names.
Chart Height Scale Factor: 1 Chart Left Margin Scale Factor: 1 Show Columns in Table: ♥ Log2FC P-Value ■FDR	only logFC

The resulting bubble plot will show all three comparisons for each gene.

CHOmics (v1) Toolbox * My Analyses	 Admini Projects 02 Companisons (16) Samples (56) 	<u> </u>		Hello, Demo User
umber of genes appeared: 7, Number of c	omparisons appeared 3 >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	Download data		
		Bubble Plot		H 0
			 D108.vs.D7 D84.vs.D7 D96.vs.D8 	16.) 1
nord14e M621d1 3832837 0689270 0689269	•	•		
Caler		•••		
-3	-2 -1 Log 2	o z 2 Fold Change	D84 vs D72 logFC	D96 vs D84 logFC
eName	* GeneIndex	D108vs D96, logFC	D84vs.D72_logFC	Dg6.vs.D84_logFC
	10029621	0.33917	130046	0.62098
	10015941	0.75588	276547	-1.00401
0689269	11010945	138505	2.31543	-1.38274
0689270	11010950	1.74787	196622	-1.48396
3832837	11004338	1.05071	1.05720	-1.11520
da	10038209	199236	238963	-1.75335
d14e	10037538	139038	2.95523	-119068

The data table below the bubble plot can also be used for filtering. Remember in the advanced settings, we choose to display logFC only, this makes it easier to look for genes that are reverted in different time points. The logFC values are colored coded (red, increase, blue, decrease), therefore we can see that most of genes show upregulation in D84, and then downregulation in D96 and then upregulation in D108.

You can also redo the plot, check all columns to include p-value and FDR in the table, and export the results to excel file.

The workflow above uses up-regulated genes as example. You can get down-regulated genes from the filter step in meta-analysis result page.

4 Visualization

4.1 Visualize Gene Expression

CHOmics provides tool to easily visualize gene expression level across multiple genes, samples and omics. For each gene, you can view its expression levels across multiple samples.

4.1.1 View Gene Expression from multiple samples

Choose the Gene Expression tool from Toolbox -> Gene Expression Plot from top menu, and enter the official symbol of genes or load gene list from saved lists. Alternatively, in the gene details page, click View Gene Expression link.

CHOmics (v1) Tootbox	Hells, Demo User 🛛 🕹 Sign Da
Gene Expression Plot Gene Expression Tool	
Genes: » Load from saved lists Q Load functional gene sets X Clear	Samples: W Load from saved lists Q Search and Select Q Select a Project X Clear
Adamta6 Adarta6 Adata Adata Aeba Arga Ampda	DK8802008 DK8802008 DK8802008 2. Enter or load sample list DK880202 DK880202 DK880202 DK80202
Note: Leave empty to view expression of all genes from selected samples.	Note: Leave empty to view expression of selected genes from all samples
Platform Type: ORVA-Seq Only, Ignore Microarray Samples: OMicroarray Only, Ignore RNA-Seq Samples: Automatic Data Type: ORVBIC only OPIvate only OAutomatic based on the entered samples. If no samples entered, will use bo	the public and private samples the samples and the the samples
Sample Attributes: (13 selected) >> Show Attributes + Filter Data by Sample Attributes	4. (Optional) Choose attributes, apply data filters
Age CetType Collection DiseaseCutegory Flag Remark Flag To Remove Infection Organism Symptom TissueCutegory Tinnifection Uberon ID Uberon Term V Other All X-Deck Nare Collect Nare V V	RIN Number 📉 RNASoq Assignment Rate 🔄 RNASoq Mapping Rater 🔅 RNASoq Total Read Count 🔄 Response 🔄 SamplePathology 📄 SamplePype
C Prot S Report All	

As an optional step, you can choose what sample attributes to pass to the plot, and use data filter to choose only a subset of data points.

The Data Filter can be very useful if there are too many data points, and you want to focus on a few diseases or tissue types.

The screenshots below show default boxplot showing all samples by different time points (i.e, treatment).

Summary of Data

16g genes found.
 15 samples found. DK1B1D108, DK1B2D108, DK1B3D108, DK1B3D108, DK1B3D108, DK1B3D72, DK1B3D72, DK1B3D72, DK1B3D72, DK1B3D84, DK1B3D84, DK1B3D84, DK1B4D84, DK1B4D84, DK1B4D96, DK1B2D96, DK1B3D96, DK1B3D96,



Customize Gene Expression Plot

The boxplot is created using CanvasXpress (<u>https://canvasxpress.org</u>)plug-in, and sample grouping and coloring can be customized by the user. In the example below, we show how data points are colored.



4.1.2 View Gene Expression in Heatmap

Heatmap can be useful to visualize gene profiles from multiple samples. It can also provide information about how genes and samples cluster.

Samples: >> Load from saved lists Q, Search and Select Q, Select a Project X Clear
DKBD04 DKBD06 DKBD06 DKBD06 DKBD06 DKBD06
rlaying on heatmap onames er Infection Organism RIN Number RNASeq Assignment Rate RNASeq Mapping Rate RNASeq Total Read Count Tarafecton © Treatment Oberon TD Uberon TD Uberon TD V Check All X Check None

You can enter genes and samples in the box, or load pre-saved genes and samples quickly from your collection. Be default, we will log2 transform the gene expression data, perform scaling of the data across samples for each gene, and limit the scaled value to -3 to 3 before displaying the data in heatmap. This works well in most situations. However, advanced users can change the options. For example, if you want to keep the order of samples as you entered, just uncheck "Cluster Samples".

Advanced Op	otions	>
Data Options		
Enable Log2 Transfo	rm	
Value To Be Added	d For Log Transformation: 0.5	
Enable Z-Score Trans	sform	
🕑 Enable Upper Limit	3	
🛛 Enable Lower Limit	-3	
Cluster Genes		
🗹 Cluster Samples		
Display Options		
Overlay Samples		
Display Gene Names	5	
Display Sample IDs		

The heatmap is rendered by CanvassXpress. You can change the plot size if needed.

Download: 📩 Raw Data File 🛓 Heatmap Data File 📕 Bookmark URL



In the example heatmap, we entered a few significantly differential expressed genes between time D72 vs time D108. From heatmap clustering, we can see that the samples are clearly clustered by time points with increase of expression on most of genes along with time.

4.1.3 Multi-omics Expression View

Besides the plotting of transcriptomics data, CHOmics also enables the visualization of other types of omics data such as proteomics, and the comparison across omics.

Here is an example of comparing gene expression (transcriptomics) and protein expression (proteomics) of gene CTSA at different time points, using the 'Gene Expression Plot' tool. By righ clicking the plotting area, users can group the samples by different treatment time points while segregating the data by omics type(i.e, Samplesource).

Hello. Demo User 🛛 😝 Sign Out

C Plot Reset All

Summary of Data

1 gene found: Ctsa
 25 samples found: DK1B1D108, DK1B2D108, DK1B3D108, DK1B4D108, DK1B1D7z, DK1B1D7z, DK1B3D7z, DK1B3D7z, DK1B3D7z, DK1B3D7z, DK1B3D84, DK1B3D84, DK1B3D84, DK1B4D84, DK1B3D96, DK1B3D96, DK1B3D96, DK1B4D96, P_DK1-B1-D108, P_DK1-B2-D108, P_DK1-B3-D108, P_DK1-B3-D108,

Download 🚣 Raw Data File



Visualize Comparison Data 4.2

Dashboard View of Comparison 4.2.1

The dashboard shows a summary of all the comparisons.



The above dashboard shows the comparisons from different Categories, Cell Type, Disease State, Treatment, Platform, etc. Below the dashboard, there is also a table listing all the comparisons.

In addition, users can set Dashboard Preference to change how the comparison summary is displayed.



4.2.2 Bubble Plot

Bubble plot is another useful demonstration of gene or gene set in comparisons. For each gene, you can view all the available comparisons in a bubble chart.

CHOmics (v1) Toolbox - My Analyse	es - Admin - Projects (1) Co	omparisons (12) Samples (30)	
Bubble Plot			
» Multiple genes .vs. multiple comparisons	Gene Symbol		
Gene Name:	Tgm2		
	Please enter the gene name,	e.g., BRWD1-IT2	
Y-axis Field:	Case_Treatment	•	
Coloring Field:	Case_SampleSource	\$	
Comparison Type:	All Comparisons	•	
	> Next Step		

The default settings work for most users. After clicking the Next Step button, you will see a plot like:



In the bubble plot, the X-axis shows log2 Fold Change of the comparison, the Y-axis shows 'Case_treatment'. Each dot represents the comparison result of this gene from one comparison. The color of the dot represent 'Case_Samplesource' (i.e, here we set as omics type), and the size of the dot represent significance (-log10(FDR), larger is more significant).

The user can click and unclick the color legend at right to select or deselect omics types. When mouse over a dot, more details are shown. And the user can also click the dot to link to other graphs.

The tool bars at top right corner allows the user to zoom and pan the graph.

The screenshot below shows the same bubble chart after selecting one omics type (i.e,transcriptomics), and zoom into a portion of the chart.



Data Filter and Advanced Settings in Bubble Plot

In addition, advanced users can change settings by click "Modify Settings Button". For example, the user may want to show a selected list of diseases. After clicking Customize in Case_Treatment, user can select which treatments to display in the pop-up window.

	Marker Area	P-Value Adjusted P-Value	Treatme	ent ¹¹ Sort by Category 11 Sort by Occurer	co	×
	Y-axis Setting	Case_Treatment Show Top 10 Show Top 20 Show All Customize		Check / Uncheck All		
	Coloring Setting			Name	Occurence	
		Show Top 10 Show Top 20 Show All Customize		D84 vs D72	2	
				Dg6 vs D7z	2	
			*	D108 vs D72	2	
				Dg6 vs D84	2	
C Modify	The plot contains t	a out of 12 data points	0	D108 vs D64	2	
Settings	The plot contains 1	2 out of 12 data points.	0	D108 vs D96	2	
🛓 Download Data						
L Download SVG						

After modifying the setting, the user can click plot button to view the new chart. The system will display how many data points are chosen based on the filter.

oad SVG			Bubb Colored b	e Chart for Tgm2 y Case_SampleSource	1	@ €+□	0.00×**
	D96 vs D72	•			•		 Transcrip Proteomia
	D84 vs D72	•	•				

Close

Bubble Plot of Multiple Genes and Multiple Comparisons

It can be useful to look at a set of genes (e.g. all differentially expressed genes, or genes from a certain pathways) in a set of related comparisons (e.g. all from the same disease).

CHOmics (V1) Toolbox • My Analyses • Admin •	Projects (2) Comparisons (16) Samples (56)
Bubble Plot	
» Multiple genes .vs. multiple comparisons	
Bubble plot of gene set Gene Name:	jak3
	Please enter the gene name, e.g., BRWD1-IT2
Y-axis Field:	Case_DiseaseState \$
Coloring Field:	Case_SampleSource \$
Comparison Type:	All Comparisons 🗢
	> Next Step

To view this type of bubble plot, select the link for Multiple Genes vs. multiple comparisons.

In the Genes and Comparisons Bubble plot window, you can now enter the symbols of the genes, and the comparison names. However, it is much easier to use the saved genes and saved comparisons features, or other tools from the system to quickly get a get set. Please see below for details.

CHOmics (v1) Toolbox • My Analyses • Admin • Projects (2) Comparisons (16) Samples (56)	Hello, Demo User	6 Sign Out
Bubble Plot Multiple		
» Single Gene Plot		
Genes: » Load from saved lists Q Load functional gene sets X Clear	Comparisons: » Load from saved lists Q Search and Select Q Select a Project X Clear	
Load saved gene list	Load saved comparisons	
Note: You must enter one or more gene names.	Note: You must enter one or more comparison names.	
02 Toggle Advanced Settings		
Submit		

In the example below, we use dashboard to select 6 comparisons that are for different time points in CHO cell lines. We save the comparisons and load in the bubble plot tool. For gene list, we get the upregulated genes from comparison D72 vs D108, and paste into the gene names fields.

In the bubble plot, the gene symbols are listed in Y-axis. The X-axis represents logFC, and color of the bubble represents comparison; the size of the bubble represents the significance.



In the legend, the color keys for comparisons are shown. You can click the color key in the legend to hide/show comparisons. The size of the color dot in the legend correlates to the largest bubble for that comparison, which is the most significant gene with the smallest FDR.

Bubble Plot of Multiple Omics data

Similar to the bubble plot of multiple genes across multiple comparisons, users can further compare the genes on the comparisons from different omics data.

Bubble Plot Multiple

ons >> Load from saved lists Q_Search and S D72 Ison_Protein_D108_D72 must enter one or more comparison names.	etect. Q Setect a Project
D72 ison_Protein_D108_D72 must enter one or more comparison names.	
must enter one or more comparison names.	
	◙ Q+⊡₽ ◘ ■X* ==
	D100 072
	Comparison_Protein_D108_D72
	2
	1 1.5

4.2.3 Get significant genes from comparisons

Another way to get a gene set to visualize in the genes/comparisons bubble plot is to filter for significantly changed genes. To do this, first select a few comparisons from the dash board, and click the "View Significantly Changed Genes" button.

	ICS (VI) HOULDARY HAY ANILYSIS - Admin	 Higects d/ Companisons d2) Sampli 	es (30)			Hello, Demo User	te sign ut
My	Private Projects						C Show
All	Comparisons Comparisons	nificant changed g	enes				S ho
Lis	t of Comparisons						BH
Save Cr	omparison List + Bubble Plot + Significantly O	hanged Genes + Pathway Heatmap + Mi	eta Analysis + Export Comparison I	Jula + WikiPuthways + Reactome	Pathways + KEGG Pathways		
		(Personality) (Resourcements)		and an			
	the second						
Gene E	Inpression Plot + Heatmap + Correlation Tool	ins				Found	
Kumn v	epression Pool + Headmap + Conveliation Tool Isability Copy CSV Show 10 \$ entr	ios				Search	
kono E lumn v	epression Pot	es v ComparisonCategory	Case_Tissue	Case_DiseaseState	Case_Treatment	Soarch: PlatformName	
ione E iumn v E	spession Plot • Heatmap • Correction Tool • Seality Copy Copy Copy Show 10 tents Name Dg6/vs.D84	ComparisonCategory Others	Case_Tissue Unknown Tissue	Case_DiseaseState Unknown Disease	Case, Treatment Dp8 vs D84	Search: PlatformName Generic Mouse NGS Platform	
ione E iumn v iei iei	eprenation Poir	ComparisonCategory Cothers Cot	Case, Tissue Unknown Tissue Unknown Tissue	Case_DiseaseState Unknown Disease Unknown Disease	Case, Treatment D96 vs D84 D96 vs D72	Search: PlatformName Generic Mouse NGS Platform Generic Mouse NGS Platform	
iumn v Lumn v K K	erenson Nict	ComparticinCalegory ComparticinCalegory Others Others Others Others Others	Case, Tissue Unknown Tissue Unknown Tissue Others	Case_DiseaseState Unknown Disease Unknown Disease Others	Case. Treatment Dg8 vs D84 Dg8 vs D72 D84 vs D72 D84 vs D72	Search: PlatformName Generic Mouse NGS Platform Generic Mouse NGS Platform Generic Mouse NGS Platform	
iumn v R R R R	spreador Nati Printenego Computation Taxi Babity Capy Cox Show 10 t entr Name Digitive Disc Digitive Disc Digitive Disc Digitive Disc Digitive Disc	ComparisonCategory ComparisonCategory Others Others Others Others	Case_Tissue Unknown Tissue Unknown Tissue Others Unknown Tissue	Case_DiseaseState Unknown Disease Unknown Disease Others Unknown Disease Unknown Disease	Case. Treatment D96 vs D84 D96 vs D72 D84 vs D72 D86 vs D95	Search:	
Kumn V Kumn V K K K K K	consolid Nati eventsory cov c	ComparisonCategory ComparisonCategory ComparisonCategory Others Others Others Others Others Others	Case_Tissue Unknown Tissue Unknown Tissue Others Linknown Tissue Unknown Tissue	Case_DiseaseState Unknown Disease Unknown Disease Others Unknown Disease Unknown Disease Unknown Disease	Case, Treatment D96 vs D64 D96 vs D72 D84 vs D72 D86 vs D72 D86 vs D72 D98 vs D84	Search: PlatformName Generic Mouse NGS Platform	
iumn v Li ium R R R R	Name Conduction Table Dg6usc D04 Dg6usc D04	ComparisonCategory ComparisonCategor	Case, Tissue Uninown Tissue Oniens Others Uninown Tissue Uninown Tissue Uninown Tissue	Case_DisesseState Unknown Disesse Unknown Disesse Others Unknown Disesse Usknown Disesse Unknown Disesse	Case, Treatment D96 vs D84 D96 vs D72 D84 vs D72 D108 vs D96 D268 vs D94 D268 vs D72	Search:	
iere E iumn v ie ie ie ie ie ie ie i	consolid National Consolid Consolid National Consolid Consoli	CompanionCalegory	Case_Tissue Urinown Tissue Others Urinown Tissue Urinown Tissue Urinown Tissue Urinown Tissue Others	Case, DiseaseState Unknown Disease Others	Case, Treatment Dg6 vs D64 Dg6 vs D72 D64 vs D72 Dr64 vs D72 Dr66 vs D66 Dr66 vs D64 Dr66 vs D64 Dr66 vs D64	Search:	
iero E iumn v ie ie ie ie ie ie ie ie ie ie ie ie ie	Computation Nation Computation Nation Copy Copy	ComparisonCalegory ComparisonCalegory Others Others Others Others Others Others Others Others Texted vs Control Texted vs Control Texted vs Control	Cate, Tissue Uninoun Tissue Others Uninoun Tissue Uninoun Tissue Uninoun Tissue Others Others Others	Case, DiseaseState Unknown Disease Unknown Disease Unknown Disease Unknown Disease Unknown Disease Unknown Disease Others Others Others	Case, Theatment Dp0 vs D84 Dp6 vs D72 DR4 vs D72 Dr08 vs Dp6 Dr08 vs D84 Dr08 vs D72 Dr08 vs D72 Dr08 vs D72 Dp10 vs D72 Dp10 vs D72	Search:	
	Comparison Protein Day	ComparisonCalegory ComparisonCalegory Others Others Others Others Others Others Others Others Treated vs Control Treated vs Control Treated vs Control Treated vs Control	Case_Tissue Unicoan Tissue Unicoan Tissue Others Unicoan Tissue Unicoan Tissue Unicoan Tissue Others Others Others	Case, DiseaseState Unknown Disease Others Others Others Others Others	Case. Theatment D99 vs 084 D96 vs 072 D84 vs 072 D164 vs 072 D164 vs 056 D166 vs 072 D166 vs 072 D166 vs 072 D96 vs 084 D96 vs 072 D96 vs 072 D84 vs 072	Search:	

Dashboard filter.

In table, select comparisons and view significantly changed genes.

In the significantly Changed Genes window, the comparisons from the previous page are already loaded. You can add or remove comparisons if needed.

Now select direction (up-, down-, or both), and use the logFC cutoff and FDR value to get a list of genes. Depending on the comparisons, sometimes you may need to adjust the logFC and FDR values to get a good list of genes. In general, for bubble plot, using <100 genes will make the graph easier to read.

Once you are happy with the gene list, you can save it. You can also export the list for later use.

CHC	Omics (v1) Too	lbox - My Analyse	es • Admin • Pro	ojects (1) Compar	risons (12) Sampl	es (30)			
Sigr	nificantly	Changed	Genes						
Compa	risons: » Load fr	om saved lists Q	Search and Select	Q Select a Proj	ect × Clear				
D108. D108. D108. D84.v D96.v D96.v	vs.D72 vs.D84 vs.D96 s.D72 s.D72 s.D84	List of select	ed comparisons						
Or, uplo	oad your compariso	on files: Choose Fil	No file chosen	» (Demo Data				
Display	Options: 🕑 Loga	FC 🗹 P.Value 🗹 F	FDR						
Fold Ch	nange Cutoff: Er	nter Value 🗢	1	Both Up- and	d Down-regulated	\$			
Statistic	c Cutoff: FDR	\$ ≤ 0.0	95 ♦			<	Increase or threshold t	decrease o select genes	
Subm Sav Sav Colum	e Comparison List e Gene List • Gen nn visibility Copy	Bubble Plot Pot CSV Show	Pathway Heatmap	Meta Analysis relation Tool F	Export Comparise CA Analysis Export Comparise	on Data Image: WikiPa xport Expression Data	thways + Reacto	me Pathways 💽 🕨	KEGG Pathways
Showin	g 1 to 10 of 15 entri	es			1				
	GeneName	Description	D108.vs.D72 - Log2FC	D108.vs.D72 - P.Value	D108.vs.D72 - FDR	D108.vs.D84 - Log2FC	D108.vs.D84 - P.Value	D108.vs.D84 - FDR	D108.vs.D96 - Log2FC
	Cd68	CD68 antigen	1.5075	0.0000	0.0000	0.9910	0.0000	0.0001	0.5591
0	Clu	clusterin	1.8516	0.0000	0.0000	12528	0.0000	0.0000	0.8301
	Ctsb	cathepsin B	1.1143	0.0000	0.0000	0.7742	0.0000	0.0000	0.4140

View Significantly Changed Genes in Bubble Plot

Back to the bubble plot, you can load the saved comparisons and saved genes and view the plot.

	Save t	o compariso	n list										
 Sav Sav Colum 	e Comparison List e Gene List • Ge nn visibilitit • Ge	Bubble Plot Pe ne Expression Plot CSV Show	thway Heatmap Heatmap + Co	Meta Analysis melation Tool F	• Export Comparise CA Analysis • Ex	on Data •. WikiPa	thways 🕨 Reacto	ome Pathways 💽 🕨	KEGG Pathways				
howin	Save to	gene list	s.D72	D108vs.D72 - PValue	D108.vs.D72 - FDR	D108.vs.D84 - Log2FC	D108.vs.D84 - P.Value	D108vsD84 - FDR	D108vs.D96 - Log2FC	D108.vs.Dg6 - PValue	D108.vs.D96 - FDR	D84vs.D72 - Log2FC	Search: D84vs.D72 - PValue
8	Cd68	CD68 antigen	1.5075	0.0000	0.0000	0.9910	0.0000	0.0001	0.5591	0.0000	0.0061	0.4969	0.0002
	Clu	clusterin	1.8516	0.0000	0.0000	12528	0.0000	0.0000	0.8301	0.0000	0.0003	0.5806	0.0000
8	Ctsb	cathepsin B	1.1143	0.0000	0.0000	0.7742	0.0000	0.0000	0.4140	0.0000	0.0006	0 3224	0.0005
	Ctsi	cathepsin L	11178	0.0000	0.0000	07777	0.0000	0.0000	0.3985	0.0000	0.0008	0 3222	0.0039
×	Dax(2	diamine oxidase- like protein 2	-1.5420	0.0000	0.0000	-0.9831	0.0000	0.0002	-0.4795	0.0002	9.0186	-0.5774	0.0000
	Gm	granulin	10717	0.0000	0.0000	0.7742	0.0000	0.0001	0.4449	0.0000	0.0049	0.2799	0.0036
	LOC100771976		1.1626	0.0000	0.0000	0.6455	0.0000	0.0002	0.2467	0.0009	0.0364	0.4993	0.0001
8	LOC103162429		12001	0.0000	0.0000	0.5994	0.0000	0.0002	0.2880	0.0003	0.0204	0.5830	0.0000
×	Mmp19	matrix metallopeptidase 19	0.9750	0,0000	0.0000	0.6372	0.0000	0.0001	0.1829	0.0014	0.0440	0,4200	0.0000
8	Mmp3	matrix metallopeptidase 3	14305	0.0000	0.0000	0,7182	0.0000	0.0001	0.2204	0.0013	0.0416	0.6943	0.0000
8	Nppb	natriuretic peptide type B	1.7812	0.0000	0.0000	0.9140	0.0000	0.0002	0.4367	0.0001	0.0100	0.8486	0.0000
8	Plat	plasminogen activator, tissue	1.0157	0.0000	0.0001	0.6146	0.0000	0.0005	0.2990	0.0005	0.0255	0.3822	0.0015

In the example below, it can be seen that most significant genes come from down-regulated direction from the first four comparisons.



4.2.4 Volcano Plot

Volcano plot is useful to view a top level summary of how many genes are significantly up- or down-regulated in a comparison.

)	2.	Search to find comparison
D108.vs.D72	Q Comparisons	
Please enter the comparison id, e.g., GS	E43696.GPL6480.test2	
P-value FDR Cutoff: 0.05	•	
2 🗘	3. (Optional)	
Volcano Chart	Change settings	
• Auto (based on cutoff) Customize		
Submit Chart Width (px): 1000	Chart Height (px): 800	» Add A New Chart
	D108vs.D72 Please enter the comparison id. e.g., GS P-value • FDR Cutoff: 0.05 2	2. D108.vs.D72 Q Comparisons Please enter the comparison id. e.g., GSE43696.GPL6480.test2 P-value • FDR Cutoff: 0.05 ÷ 2 ÷ 3. (Optional) Change settings • Auto (based on cutoff) Customize Submit Chart Width (px): 1000 Chart Height (px): 800



You can use mouse to drag over an area to zoom in.

Mouse over a point will show the gene details. Click the data point will show you links to other graphs.

View Multiple Volcano Plots Together

Users can also show multiple comparisons side-by-side. If needed, the user can also highlight the same group of genes across the volcano plots.

Comparison Name:	D108.vs.D72	Q Comparisons	The first
	Please enter the comparisor	id. e.g., GSE43696.GPL6480.testz	comparison
Y-axis Statistics:	P-value FDR Cuto	ff: 0.05 ¢	
Fold Change Cutoff:	2 🗘		
Chart Name	Volcano Chart		
Comparison ID	D108.vs.D84	Q Comparisons	The second
	Please enter the comparisor	n id, e.g., GSE43696.GPL6480.test2	comparison
Y-axis Statistics:	OP-value OFDR		
Chart Name	Volcano Chart		
Fold Change Cutoff:	2 💠	Use "Customize" to	highlight
Statistic Cutoff:	0.05	genes entered in the	text box

The resulting volcano plots are shown as below. Selected genes are shown as orange dots.



4.3 Visualize functional pathway

4.3.1 Enrichment from Up and Down Regulated Genes

When you view details of a comparison, the functional enrichment results are shown. Briefly, for each comparison, we generated the up- and down- regulate gene lists, and use these lists to compare with all genes in the genome to identify functions that are significantly enriched.

CHOMICS (VI) Toobor • My A	Valyes.+ Adver+ Prosterio Companyorititi	Rempton (pt)						Heliz Dens User 39 Sign Dut
Comparison: D108.VS	D72							
ID Name	Project		Category	DiseaseState	Toour	Contrast	Case Samples	Control Samples
3 0506vs.D72	Child Dente - Child Dente RNA-Sep Analysis			Unknown Disease	Unknown Tissue		Show/Hole	Show/ride
O Ves Delats Green Charged	Cones Mildfordsongs Beachama Indhongs #200	Patricia Patricia India Analysia	is Votano Chart					
Up 5	legulated Genes	· Obertaad SVG Rev			Top10 pat	hways enric	hed	
Notopical Process			KEG	G				
Cellular Cumponent					10000-04-10			
Motecular Function	D.G. Jaimle		al advestor		ng(pt-15.80			
x600	Wuttiple	e patriway uatabases	rinteraction	inggi 12.54				
Molecular Signature			Necetore	ing25-11.47				
Planto Procen Contrain			HAPK algoaling pathway	ME20-10.00	NIGOD 48-44			
Beactorie			The expected pathway (34 Akt eignaling pathway	1022-424	10ggs1-8.13			
I Ereichment Report				20 et	61			
Down	Regulated Genes	In Downson SVG File						
Bological Process			KEG	G				
Cellubr Component			2003-000-	Name and American				
Hotecular Function			DAA replication Call cycle		1000 Magazi - 47, 52			
4800			Ribosome Mismatch repair	Mg2(-47.31	Ingelas - 19. 73			
Holecular Signature			Fances anemia pattway	Vigos-24-38				
Printpro Problen Domain		Proyetterune-tre	Base excelor repair	Register-13.82				
Wild Pathwity			Oscyte metabolism Pyrimidine metabolism	10000-12.14 10000-12.40				
Reactorner			•	10 20 30	40 30			
# Liveryward slepost				Number of Genes				
	View full repor	t						

In the example above, this comparison is between D108 vs D72, and the top up-regulated biological processed are response to virus, immune effector process.

Click the left menu will switch the bar charts for different categories (Gene Ontology, KEGG, Molecular signature, Protein domain etc).

The bar charts here show the top 10 categories. To view complete results, click the Enrichment Report.

Inriche	d Categories						Top10 pathw	ays enriche	d	
oggie colum	ins 👩 GO Thee 👩 TermiD 👩 Term 👩 Enrichment	B logP B Function	nal Terms	nget	Genes 👩 Total Gen	os	1			
Column visib	Ety Copy CSV Show 10 \$ entries						-	Search	-	
GO Tree :	TermID	Term	Enrichment *	logP	Genes in Term	Target Genes in Term	Fraction of Targets in Term	Total Target Genes	Total Genes	Action
MSigD8	LEE_BMP2_TARGETS_DN	LEE_BMP2_TARGETS_DN	3.91763224871508e- 31	-70.0146504296904	807	153	0.13871260199456	1103	15720	EShow Genes
Gene Ontology	GO:0044424	intracellular part	3 36305799347635e- 25	-58.3517766476754	13240	933	0.75980198019802	1212	20830	EShow Genes
Gene Ontology	GO:0005622	intracellular	6.67887260173594e- 25	-55.6656781237414	13317	936	0.772277227722772	1212	20830	EShow Genes
Gene Ontology	GO:0005488	binding	9.43954474027105e- 25	-65.3197195725229	12660	892	0.758503401360544	1176	20347	EShow Genes
Gene Ontology	GO 0005730	nucleolus	5.912051254420940+ 24	-53.4850493786918	813	125	0.103135313531353	1212	20830	IEShow Genes
Gene Ontology	GD 0044464	cell part	52815457378631e- 23	-512952383306056	15460	1037	0.855510581056106	1212	20830	EShow Genes
Gene Ontology	GO:000558:3	cell	6.336974990944028- 23	-511127400568371	15465	1037	0.855510561056106	1212	20830	IEShow Genes
MSigDB	GSE6674_ANTI_IGM_VS_CPG_STIM_BCELL_DN	GSE6674_ANTI_IGM_VS_CPG_STIM_BCELL_DN	1.53486443903669e- 21	-479258448890668	186	56	0 05077062555566364	1103	15720	Eshow Genes
MSigDB	KRIGE_RESPONSE_TO_TOSEDOSTAT_24HR_DN	KRIGE_RESPONSE_TO_TOSEDOSTAT_24HR_DN	1.614456823888970- 21	-47.8752883847299	865	140	0.126926563916591	1103	15720	EShow Genes
Gene Ontology	G0:0043229	intracellular organelle	5.911950343418620- 21	-46.5773111685508	11590	829	0.683993399339934	1212	20830	Eshow Genes

In the enrichment report, the full list of functional terms are shown by order of p-value.

4.3.2 View Changed Genes from a Functional Term in Volcano Plot

From the bar chat, click a functional term, and you have the option to view these genes in a volcano plot.

			2.View	selected pathv	vay in volcano pl	ot		
CHOmics (v1) Toolbo	ac≁ MyAninyas.* Altrin.*	Priporta III - Companiore Riff - Samples	Lysosome:		.*			Hello, Demo deer 🗰 Sign Dut
Comparison: D10	8.VS.D72 View Companion Genes *	Edit Comparison Details	Volcano Plot KEGC Pathway Save Gene List					
ID Name	Project				OK	Contrast	Case Samples	Control Samples
3 D10Rvis.D7z	CHD Demio - CHO Dem	g RNA-Sing Analysis		Unknown Disease	Unknown Tissue		Show/Hide	Shaw/hide
Up Reg	gulated Genes	>> Download SVG Fee		KEGG	1.Click to	select a pa	thway to viev	v
Biological Process				KEGG			-	
Cellular Component								
Moleciular Function				Focal adhesion	- Charles	-18.60		
KEGO			ECM-rece	ptor interaction	Aug 21-12-20	10.000		
Molecular Signature			Auto	Apoptosus	Region - 10.52			
Interpro Protéin Domain			Rheu MAPK sig	matold arthritis	No(5)-10.00	impton -6 Ab		
Wilks Fishtneny			TN# sig	naling pathway	hig(p)-6.28			
Reactome			HISK-AKE DO	0	20 40	50		
>> Encicitation Report					Number of Genes			

Once you click the link in the popup window, volcano plot will be generated for the comparison with the changed genes from the selected term highlighted.



4.3.3 View Enriched Pathways Directly from Comparison Details

From the bar chat, if you are viewing KEGG or wikipathway database, clicking the pathway name and you have the option to view pathway plot.

			3.Click to	view the pat	hway				
CHOmics (v1)	Toolbox * My Analyses * Admin * Pri	poctalli - Companioni GE - Samplità	Lysosome					Hillo, Demo Uver	Stan Dut
Comparison: E * Search All Company	0108.vs.D72 ons	t Compension Details	Voicano Piot KEGC Pathway Save Gene List						
ID Name	Project				OK	Contrast	Case Samples	Control Samples	
3 D108v1.D7	z OHD Demo - CHO Demo R	NA-Sing Analysis	Unix	noan Disease	Unknown Tissue		Show/Hide	Show/hide	
O Vew Details	mes Changed Denne WitePathways	Teactorie Pathaoya REGI Pathaoga	Pattiway Hastmap Thursde Plat M	ta Analysia Vojcana Chart					
	ip Regulated Genes	* Download SVG File			2.Click to se	lect a path	way to view		
Hiological Process	1 select a nathway	v database to view		KEGG					
Celtulat Component		y database to view				morpi -44.00			
Molecular Function			Frical	adhesion 0	- National Contraction	-19.50			
KEOO			Protooglycans ECM-receptor In	n center in	Registi-to be	10.68			
Molecular Signature	5/		Autophagy	- animai	Kog(0) - 11 64				
Interpro Protein Do	TIN		Rheumatoid	arthritia	NUCCE-10.00				
Wike Platframy			MAPK signaling The signaling	pathway pathway	Aug(a) - 6.28	ingto: -6 AB			
Conctome.			VT3K-Akt signaling	Destroay	20 40	kap(m-6.53			
>> Envictment Hope	at				Number of Genes				

This will automatically open the pathway visualization page, and preload the pathway and comparison. Click submit to view the pathway.



Gene Set Enrichment from Ranked Genes

For each comparison, we produce a rank file for all genes using logFC. We use PAGE (Parametric Analysis of Gene Set Enrichment) to identify significant biological changes. PAGE can be more sensitive for comparisons where the logFC is relatively small, but most genes in a functional set show the same direction of change.

The predefined gene sets were from MSigDB.

For each comparison, the top up-regulated and down-regulated gene sets are plotted.

PAGE Report for Up-Regulated Games. P Download SVG Kan				Check	gene s	et	BACE Report for Down Reputated Germs								
	PAGE F	Not - Up-regu	lated Genes					PAGE P	Not - Down	-regulated	Genes				
COLLACEN CONTINUES EXTRACTLUCAR MUTRIX OCCC 00 E. INITIORIA, COMONIENT OF RADIMA REMARKARE COCC DO E005. INANA COLE MUTRICOM REMODIO E ANARC COME NATURISME EXTRACTLUCAR MUTRIX DEGANIZATIONE RADICTORE E HIRA 1. INITIONES COMPONENT OF FLAMMA REMARKE DOCC DO E01.	•	•	Nama NTEORU Total Genes: 472 2 Socie: 12.87 Prvsuke: 0.0007138	L COMPONENT_OF	e jelasma membr	WWE_GODC_00_00	DAV. DEPENDENT DVA. HEILLICKTEIN ODER OG BORSHET ORIENDENDEN, HATT DOCC DE BISHET HALLIMARE (21M O-BOODENT MISSIDE C) HULLIMARE (21M O-B- CELL OVLE OHONOTET MISSIDE C) HULLIMARE (21M O-B- CELL OVLE OHONOTET MISSIDE C) HULLIMARE (21M O-B- CHARDENDENT EDCC DE BISHET		Do	ot size gnifica	repre ince (F	esent FDR)	s		
EXTRACELLUAR MATRIX GOC DO 001012 EXTRACELLUAR MEDION MATTOCCC DO 001012 EXTRACELLUAR MEDION MARCE DOC DO 0009415 EXTRACELLUAR MEDION GOCC DO 0009515 NARA HATRISONE MESIGIN CO NARA KATRISONE				ĺ	•		HITOTIC CELL CYCLE COMP GO BODEZTA CELL CYCLE MITUTIC REACTORE & HAR 48278 3 CELL CYCLE REACTORE DATABASE ID HELARGE 46 1641170 HALIMARK EIN TANGETS HESODE C2 HALIMARK E27 TANGET.	•	:					•	
	11	12	13 Z	14 Score	15	16	-11	2.5 -12	-11.5	-11	-10.5 Z Score	-10	-9.5	-4	-8.5

To view the full list of gene sets, you can click the report for genes as shown in following figure.

Column vability Copy CSV Show 10 C entries		Search					
Gene Sot Name	* # Genes	Z Score	P Value	FDR			
A, TETRASACCHARIDE, LINKER, SEQUENCE, IS, REQUIRED, FOR, GAG, SYNTHESIS, REACTOME, R, HSA, 197475, 1	15	35026	0 000999	0.008753			
ACTIN_BASED_CELL_PROJECTION_GOCC_GO_6098858	123	3.8517	0.000999	0.008753			
ACTIN_BINDING_GOMF_GO_0003779	256	41771	0.000000	0.008753			
ACTIN, FILAMENT, BASED, PROCESS, GOBP, GO, 0030029	349	3499	0.000999	0.008753			
ACTION, POTENTIAL, GOBP, GO, 0001508	36	57785	0.000999	0.008753			
ACTIVATION_OF_IMMUNE_RESPONSE_GOBP_GO_0002253	127	36063	0.000999	0.008753			
ACTIVATION_OF_MATRIX_METALLOPROTEINASES_REACTOME_DATABASE_ID_RELEASE_66_1592389	15	88731	0.000999	0.008753			
ACTIVE_TRANSMEMBRANE_TRANSPORTER_ACTIVITY_GOMF_GO_0022804	174	5 2265	0.000309	0.0087535			
ADAPTIVE_IMMUNE_RESPONSE_BASED_ON_SOMATIC_RECOMBINATION_OF_IMMUNE_RECEPTORS_BUILT_FROM_IMMUNOGLOBULIN_SUPERFAMILY_DOMAINS_GOBP_GO_0002460	83	3:3924	0.000999	0.0087535			
ADAPTIVE_IMMUNE_RESPONSE_GOBP_GO_0002250	124	37329	0.000999	0.008753			

UP-Regulated PAGE Report Details

4.3.4 Multi-layer visualization

If you are interested in a particular pathway, sometimes it is useful to map the RNA-Seq or microarray data to the pathway for visualization.





In the pathway plot, typically we use red-blue color scale to show the log2 Fold Change. Blue is down-regulated, red is up-regulated.

Pathway Plot from Several Comparisons

The user can add multiple comparisons from the pathway plot tool by clicking Add Comparison link. Besides showing log2 Fold Change, the user can also show statistical significance by clicking Enable Second Visualization Columns.

CHOmics (v1) Toolbox - My Analyses - Admin - Projects (1) Comparisons (12) Samples (30)
KEGG Pathway View
Start Over Note: * denotes required fields.
Glycolysis / Gluconeogenesis 🛓 Download Pathway File
>> Select Pathway
Comparisons: >> Load from saved lists Q Search and Select Q Select a Project × Clear
D84.vs.D72 D108.vs.D72 Choose multiple comparison
Or, upload your comparison files: Choose File No file chosen >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
Gradient Blue-White-Red (-1,0,1)
Submit

The pathway plot will now have multiple color bars corresponding to the different comparisons.



4.3.5 Pathway Heatmap From Comparisons

Users can display the enriched pathways from several related comparisons, and visualize the top enriched pathways across comparisons. Users can mix public data and inhouse comparisons.



The heatmap shows pathways in rows, comparisons in columns. The statistical significance is colorcoded (log P-value, or Z-score). Pathways are sorted by the negative logP values from the highest to the lowest.



From the pathway heatmap, users can click any data point to view details.

way Names and GeneSets: Regulated Genesets						D	View Compariso	n Detail	*
anscriptional misregulation in cancer uman papitlomavirus infection uid shear stress and atherosclerosis 24-Adt signaling pathway									Close
VF signaling pathway APK signaling pathway slactose metabolism neumatoid arthritis otein processing in endoplasmic reticulum									Fructose and mannose metabolism Progesterone-mediated oocyte maturati Huntington's disease Base excision repair Alzheimer's disease
Draw Heatmap									
Jp-Regulated Genesets vs. Comparisons, log	10(p-valu	e) 🛓 Sav	e SVG						
	D84,vs.D72	D96.vs.D72	D108.vs.D72	D96.vs.D84	D108.vs.D84	D108.vs.D96			
Lysosome HIF-1 signaling pathway Focal adhesion Ribosome biogenesis in eukaryotes FoxO signaling pathway Proteoglycans in cancer	-			x: D108.v y: Lysoso z: 44.0820 logP: -44, number o	s.D72 me 867304813 0828673048 f genes: 47	135	10 8		
ECM-receptor interaction							6		
ELM-receptor interaction Axon guidance Apoptosis Autophagy - animal Sphingolipid metabolism Protein processing in endoplasmic reticulum Recumatioid arthritis Caladeae motabolism							4		

5 Customized analysis pipeline

5.1 Use alternative tool or algorithm

The analysis pipeline is modular, each step can be modified by uses to use an alternative method if desired. The users should be familiar with the Linux bash to run the analysis steps and be familiar with php programming to make modification to the source code.

The full analysis pipeline has four steps, and each step is listed in a bash file in the analysis folder in the system.

- step_0.sh FASTQC of raw data
- step_1.sh Alignment to genome
- step_2.sh Gene count
- step_3.sh DEG detection and functional enrichment

These bash files are created by PHP programs chomics/app/bxgenomics/bxgenomics_exe_analysis.php, when users launch analysis pipeline online in a web browser via chomics/app/bxgenomics/analysis.php. For example, the current pipeline uses subread to perform alignment. If users want to modify the pipeline to change it to use the STAR program for alignment, they need the following steps:

1) Install STAR program on the server, prepare STAR index for the CHO genome.

2) Check the commands in step_1.sh, and change the commands as needed. In this case, the subread command (subjunc step) needs to be replaced by the equivalent STAR command. Since STAR can sort the bam files, the samtools sort step can be omitted. Finally, the STAR output file is named as SampleIDAligned.sortedByCoord.out.bam, an extra step is needed to rename it to SampleID.sorted.bam, so step2.sh can output gene count files with the correct sample names.

3) Edit PHP program chomics/app/bxgenomics/bxgenomics_exe_analysis.php, find the part that generates step_1.sh (The section is marked as "Step 1. Alignment with Subread"), and then make changes accordingly.

4) Test the updated system to make sure it works as expected.